MODERN ALGEBRA

By B. L. VAN DER WAERDEN, Ph.D.
PROFESSOR OF MATHEMATICS
AT THE UNIVERSITY OF AMSTERDAM

In part a development from lectures
By E. ARTIN and E. NOETHER

VOLUME I

Translated from the second revised German edition
By FRED BLUM

With revisions and additions by the author

1949

PREFACE TO THE SECOND EDITION

When revising the first volume, I took pains to bring the book up to date. To achieve this goal, it was first of all necessary to treat the fundamentals of valuation theory more thoroughly and in greater detail. I am indebted to Dr. M. Deuring for his aid in the preparation of the new chapter on *Fields with a Valuation.*

Secondly, without losing sight of the main purpose of the book, it proved possible to transform, for beginning graduate students, the first volume into a complete textbook on higher algebra, except for the theory of determinants. For this reason, Euler's theory of resultants, as well as the theory of linear equations was taken over from the second volume into the first, a section on partial fraction decomposition was added, the theory of differentiation as well as the interpolation theory were further developed, the theory of factorization was treated in a more elementary manner, and many details were made easier. The concepts of vector space and hypercomplex systems are introduced at an earlier stage, and some fundamental theorems on dimensions, norms, and traces are now proved generally (rather than only for commutative fields, as was done before).

Thirdly, I have tried to avoid as much as possible any questionable set-theoretical reasoning in algebra. Unfortunately, a completely finite presentation of algebra, avoiding all non-constructive existence proofs, is not possible without great sacrifices. Essential parts of the algebra would have to be eliminated, or the theorems would have to be formulated with so many restrictions that the text would become unpalatable and certainly useless for a beginner. On the other hand, it was possible at least to compile the building stones for a constructive foundation of algebra insofar as they exist at this time. In the theory of fields I did this by presenting the field-theoretical operations in a finite number of steps in such fashion that the intuitionistic foundation of the theory, insofar as it is possible today, can be seen readily. The theory of factorization is likewise presented in a more finite manner.

With the above mentioned aim in mind, I completely omitted those parts of the theory of fields which rest on the axiom of choice and the well-ordering theorem. Other reasons for this omission were the fact that, by the well-ordering principle, an extraneous element is introduced into algebra and, furthermore, the consideration that for virtually all applications the special case of countable fields, in which the counting replaces the well-ordering, is wholly sufficient. The beauty of the basic ideas

of Steinitz' classical treatise on the algebraic theory of fields is plainly exhibited in the countable case.

By omitting the well-ordering principle, it was possible to retain nearly the original size of the book in spite of the above mentioned supplements.

At this point, I should like to express my heartfelt appreciation to a number of friends who, by their constructive criticism, contributed many minor improvements, mostly of a didactic nature. I hope that the more elementary and in some parts more detailed exposition of the content of this book has increased its usefulness, especially for beginners.

B. L. VAN DER WAERDEN

Leipzig, January 1937

PREFACE TO THE ENGLISH EDITION

The present English translation has given me a welcome opportunity for several minor changes. Following a suggestion made by Zariski in his review of the second edition, I have restored the original way of introducing polynomials independent of hypercomplex number systems. At the end of the first volume I have treated in greater detail the valuations of algebraic number fields. The original treatment was complete, but certainly too condensed.

B. L. VAN DER WAERDEN

Baltimore, May 1948

CONTENTS

PAGE

INTRODUCTION ix

CHAPTER I

NUMBERS AND SETS

1. Sets 1
2. Mappings. Cardinality 2
3. The Number Sequence 3
4. Finite and Countable (Denumerable) Sets 7
5. Partitions 9

CHAPTER II

GROUPS

6. The Group Concept 11
7. Subgroups 18
8. Complexes. Cosets 22
9. Isomorphisms and Automorphisms 24
10. Homomorphisms. Normal Divisors. Factor Groups 27

CHAPTER III

RINGS AND FIELDS

11. Rings 32
12. Homomorphism and Isomorphism 38
13. The Concept of a Field of Quotients 39
14. Vector Spaces and Hypercomplex Systems 42
15. Polynomial Rings 45
16. Ideals. Residue Class Rings 49
17. Divisibility. Prime Ideals 53
18. Euclidean Rings and Principal Ideal Rings 55
19. Factorization 58

CHAPTER IV

POLYNOMIALS

20. Differentiation 63
21. The Zeros of a Polynomial 64

PAGE
22. Interpolation Formulae 66
23. Factorization 70
24. Irreducibility Criteria 74
25. Factorization in a Finite Number of Steps 77
26. Symmetric Functions 78
27. The Resultant of Two Polynomials 83
28. The Resultant as a Symmetric Function of th Roots. 86
29. Partial Fraction Decomposition 88

CHAPTER V

THEORY OF FIELDS

30. Subfields. Prime Fields 91
31. Adjunction 93
32. Simple Field Extensions 94
33. Linear Dependence over a Skew Field 99
34. Linear Equations over a Skew Field 104
35. Algebraic Field Extensions 106
36. Roots of Unity 111
37. Galois Fields (Finite Commutative Fields) 115
38. Separable and Inseparable Extensions 119
39. Perfect and Imperfect Fields. Root Fields 124
40. Simplicity of Algebraic Extensions. Theorem of the Primitive Element 126
41. Norms and Traces 128
42. The Field-theoretical Operations in a Finite Number of Steps 134

CHAPTER VI

CONTINUATION OF THE GROUP THEORY

43. Groups with Operators 138
44. Operator Isomorphism and Operator Homomorphism 140
45. The Two Laws of Isomorphism 141
46. Normal Series and Composition Series 142
47. Direct Products 146
48. Simplicity of the Alternating Group 149
49. Transitivity and Primitivity 150

CHAPTER VII

THE GALOIS THEORY

50. The Galois Group 153
51. The Fundamental Theorem of the Calois Theory 156
52. Conjugate Groups, Conjugate Fields and Elements 159
53. Cyclotomic Fields 160
54. The Periods of the Cyclotomic Equation 163

PAGE
55. Cyclic Fields and Pure Equations 168
56. Solution of Equations by Radicals 172
57. The Generic Equation of Degree n 175
58. Equations of the Second, Third, and Fourth Degrees 177
59. Constructions with Ruler and Compass 183
60. Metacyclic Equations of Prime Degree 187
61. Establishment of the Galois Group. Equations with a Symmetric Group 189

CHAPTER VIII

INFINITE FIELD EXTENSIONS

62. Algebraically Closed Fields 193
63. Simple Transcendental Extensions 197
64. The Degree of Transcendence 200
65. Differentiation of Algebraic Functions 202

CHAPTER IX

REAL FIELDS

66. Ordered Fields 209
67. Definition of the Real Numbers 211
68. Zeros of Real Functions 218
69. The Field of Complex Numbers 222
70. Algebraic Theory of Real Fields 225
71. Existence Theorems for Formally-Real Fields 229
72. Sums of Squares 233

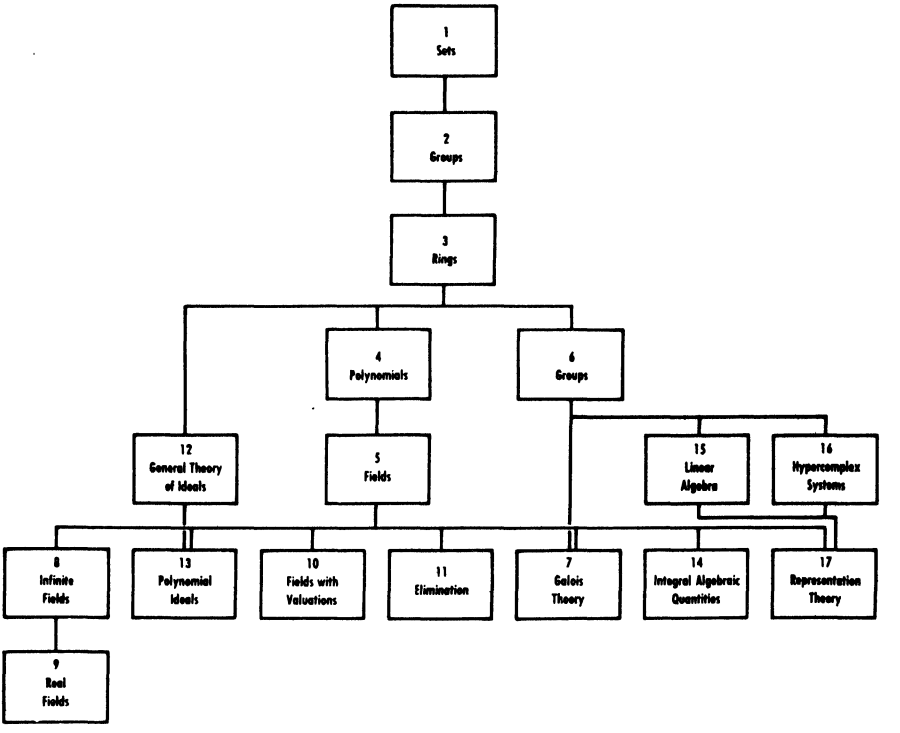
CHAPTER X

FIELDS WITH VALUATIONS

73. Valuations 235
74. Complete Field Extensions 240
75. Valuations of the Field of Rational Numbers 244
76. Valuation of Algebraic Extension Fields 248
77. Valuations of Algebraic Number Fields 255
INDEX 259

GUIDE

Logical interdependence of the chapters of volumes I and II:



INTRODUCTION

PURPOSE OF THIS BOOK. The recent expansion of algebra far beyond its former bounds is mainly due to the "abstract," "formal," or "axiomatic" school. This school has created a number of novel concepts, revealed hitherto unknown interrelations, and led to far-reaching results, especially in the theories of *fields* and *ideals*, of *groups*, and *hypercomplex numbers*. The chief purpose of this book is to introduce the reader into this whole world of concepts. Within the scope of these modern ideas classical results and methods will find their due place.

DISTRIBUTION OF SUBJECT MATTER. DIRECTIONS FOR THE READER. In order to develop with sufficient clarity the general viewpoints which dominate the "abstract" conception of algebra, it was necessary to present afresh the fundamentals of group theory and of elementary algebra.

In view of the fact that competent expositions on group theory, classical algebra, and the theory of fields have been published recently,[1] it was possible to present these introductory chapters briefly (but completely).

I have tried to write this book in such a way that each part of it should be intelligible by itself, as far as possible. Those who wish to become acquainted with the general theory of ideals or with the theory of hypercomplex numbers need not study Galois theory before, and vice versa; and those who want to consult the book about elimination of linear algebra need not be deterred by complicated ideal-theoretical terms.

For this reason the subject matter has been distributed in such fashion that the first three chapters contain a most concise exposition of what is prerequisite to *all* subsequent chapters: The fundamentals of the theories of 1. sets, 2. groups, 3. rings, ideals, and fields. The remaining chapters of the first volume are in the main devoted to the theory of commutative fields and are based primarily on Steinitz'

---

[1] For group theory the reader is referred to:

Speiser, A.: *Die Theorie der Gruppen von endlicher Ordnung*, 2nd edition, Berlin. Julius Springer 1927.

For the theory of fields:

Hasse, H.: *Höhere Algebra I, II* and *Aufgabensammlung zur Höheren Algebra.* Sammlung Göschen 1926/27.

For classical algebra:

Perron, O.: *Algebra I, II.* 1927.

For linear algebra:

Bôcher, M.: *Introduction to Higher Algebra.* New York 1908.

fundamental treatise in *Crelles Journal*, Vol. 137 (1910).[2] The theory of modules, rings, and ideals with applications to elimination theory, elementary divisors, hypercomplex numbers, and group representations will be treated in the second volume in several, mostly independent chapters.

The theory of algebraic functions and the theory of continuous groups had to be omitted, since an appropriate treatment of both involves transcendental concepts and methods. Because of its extent, the theory of invariants could not be included, either. Determinants are assumed to be known.

For further information we refer the reader to the table of contents, and especially to the foregoing schematic diagram which illustrates exactly how many of the preceding chapters are requisite to each of the chapters.

Less important or more difficult addenda are printed in small type.

The last three chapters of the first volume may be omitted on first reading.

The interspersed *exercises* may serve as a test whether the subject has become clear to the reader. Some of them contain examples and supplements, which are sometimes referred to in later chapters. No special devices are necessary for their solutions unless indicated in square brackets.

Sources: This book has in part grown out of the following courses:

Lectures given by E. Artin on *Algebra* (Hamburg, Summer session 1926).

A seminar on *Theory of Ideals*, conducted by E. Artin, W. Blaschke, O. Schreier, and the author (Hamburg, Winter 1926/27).

Lectures by E. Noether, both on *Group Theory* and *Hypercomplex Numbers* (Göttingen, Winter 1924/25 and Winter 1927/28).[3]

New proofs or new arrangements of proofs in this book are in most cases due to the lectures and seminars mentioned, regardless of whether or not the source is expressly quoted.

---

[2] This treatise has meanwhile been published in book form. Steinitz, E.: *Algebraische Theorie der Körper mit Erläuterungen und einem Anhang von R. Baer und H. Hasse.* Berlin 1930.

[3] An elaborate treatment of the latter course by E. Noether was published in *Math. Zeitschrift Vol. 30 (1929)* pp. 641-692.

CHAPTER I

NUMBERS AND SETS

We begin with a brief chapter on certain logical and general mathematical concepts, especially that of a set, which are used throughout this book, but are often unfamiliar to a beginner. We shall not, however, be concerned with basic points of difficulty.[1] While we shall always assume a "naive viewpoint," at the same time we shall avoid definitions which contain a vicious circle and thus lead to paradoxes. The more advanced student need remember only the meaning of the symbols $\in$, $\subset$, $\supset$, $\cap$, $\vee$ and $\{..\}$, as explained in this chapter, and may skip the rest of it.

1. SETS

As a starting point for all mathematical considerations we take certain objects, such as numerals, letters, or their combinations. A *set* or a *class* is defined by any property which any single one of these objects does or does not have. Those objects which have this property are called *elements of the set.* The symbol

$$a \in \mathfrak{M}$$

means: a is an element of $\mathfrak{M}$; geometrically speaking we say: a lies in $\mathfrak{M}$. A set will be called empty if it does not contain any elements.

We assume that it is legitimate to regard sequences and sets of numbers again as objects and elements of sets (sets of second order, as they are sometimes called). These sets of second order may again be elements of sets of a higher order, etc. However, we shall avoid such terms as "the sets of all sets," since they may give rise to contradictions (and have done so in the past); we shall rather form new sets only from a previously strictly defined category of objects (to which the new sets do not belong).

If all elements of a set $\mathfrak{N}$ are at the same time elements of $\mathfrak{M}$, then $\mathfrak{N}$ is called a *subset* (*subclass*) of $\mathfrak{M}$ and we write:

$$\mathfrak{N} \subseteq \mathfrak{M}.$$

Also $\mathfrak{M}$ is said to *include* or *comprehend* $\mathfrak{N}$ and we write:

$$\mathfrak{M} \supseteq \mathfrak{N}.$$

If $\mathfrak{A} \subseteq \mathfrak{B}$ and $\mathfrak{B} \subseteq \mathfrak{C}$, then $\mathfrak{A} \subseteq \mathfrak{C}$.

The empty set is contained in every set.

[1] Cf. A. Fraenkel: *Einführung in die Mengenlehre*, 3rd edition (Berlin 1928).

If all elements of $\mathfrak{M}$ are in $\mathfrak{N}$, and all elements of $\mathfrak{N}$ are in $\mathfrak{M}$, then the sets $\mathfrak{M}$ and $\mathfrak{N}$ are said to *coincide* or to be *equal:*

$$\mathfrak{M}=\mathfrak{N}.$$

Thus the equality $\mathfrak{M}=\mathfrak{N}$ means that the relations

$$\mathfrak{M}\subseteq\mathfrak{N}, \qquad \mathfrak{N}\subseteq\mathfrak{M}$$

hold simultaneously. Or we may say: Two sets are equal if they contain the same elements.

If $\mathfrak{N}\subseteq\mathfrak{M}$ without being equal to $\mathfrak{M}$, $\mathfrak{N}$ is called a *proper subset* of $\mathfrak{M}$, or we say $\mathfrak{M}$ contains $\mathfrak{N}$ *properly*. We write

$$\mathfrak{N}\subset\mathfrak{M}, \qquad \mathfrak{M}\supset\mathfrak{N}.$$

Thus, $\mathfrak{N}\subset\mathfrak{M}$ means that all elements of $\mathfrak{N}$ are in $\mathfrak{M}$ and that there is at least one more element in $\mathfrak{M}$ not belonging to $\mathfrak{N}$.

Let $\mathfrak{A}$ and $\mathfrak{B}$ be two arbitrary sets. The set $\mathfrak{D}$ consisting of all elements common to $\mathfrak{A}$ and $\mathfrak{B}$ is known as the *intersection* of $\mathfrak{A}$ and $\mathfrak{B}$ and we write:

$$\mathfrak{D}=[\mathfrak{A},\mathfrak{B}]=\mathfrak{A}\cap\mathfrak{B}.$$

$\mathfrak{D}$ is a subset of both $\mathfrak{A}$ and $\mathfrak{B}$. Any set having this property is contained in $\mathfrak{D}$.

The set $\mathfrak{B}$ consisting of all elements that belong to at least one of the sets $\mathfrak{A}$ and $\mathfrak{B}$ is called the *union* of $\mathfrak{A}$ and $\mathfrak{B}$:

$$\mathfrak{B}=\mathfrak{A}\vee\mathfrak{B}.$$

$\mathfrak{B}$ includes both $\mathfrak{A}$ and $\mathfrak{B}$, and any set including $\mathfrak{A}$ and $\mathfrak{B}$ includes $\mathfrak{B}$ as well.

The same definition holds for the intersection and union of an arbitrary set Σ of the sets $\mathfrak{A}, \mathfrak{B}, \ldots\ldots$. For the intersection (i.e. the set of the elements which lie in all sets $\mathfrak{A}, \mathfrak{B}, \ldots\ldots$ of the set Σ) we write:

$$\mathfrak{D}(\Sigma)=[\mathfrak{A},\mathfrak{B},\ldots].$$

Two sets are said to be *mutually exclusive* or *disjoint* if their intersection is empty, i.e. if the two sets have no common elements.

If a set is given by the enumeration of its elements, e.g., if we say, the set $\mathfrak{M}$ is to consist of the elements a, b, c, we write:

$$\mathfrak{M}=\{a, b, c\}.$$

This notation is justified, since, according to the definition of equality of sets, a set is determined by its elements. The defining property characterizing the elements of $\mathfrak{M}$ is: to be identical with a or b or c.

2. MAPPINGS. CARDINALITY[2]

If, by any given rule, there corresponds to each element a of a set $\mathfrak{M}$ a single new object $\varphi(a)$, then we shall call this correspondence a function, and the set $\mathfrak{M}$ is

[2] Translator's note: This term is used by Edward Kasner (cf. his book *Mathematics and Imagination*) instead of the less elegant "cardinal number," which is frequently employed. Some authors use "power" or "potency" (Bochner in his lectures on commutative algebra), both being a moral literal translation of Cantor's *Mächtigkeit*. Sets having the same cardinality might well be called *equipotent* (*gleichmächtig*).

called the *domain of definition* of the function. The set $\mathfrak{N}$ of all function values $\varphi(a)$ is called the *range of values* of the function. Such a correspondence, where to each element of $\mathfrak{M}$ there corresponds exactly one element of $\mathfrak{N}$, and where all elements of $\mathfrak{N}$ are used at least once, is also known as a (unique) *mapping* of the set $\mathfrak{M}$ upon the set $\mathfrak{N}$. The element $\varphi(a)$ is called the *image* of a, while a is an *inverse image* (*pre-image*) of $\varphi(a)$. The image $\varphi(a)$ is uniquely determined by a but the converse does not necessarily hold. Throughout the entire book the term mapping will exclusively be used for these unique mappings.

If each element of $\mathfrak{N}$ appears only once as an image, then the mapping is called *biunique*, or we refer to it as a *one-to-one correspondence*. In this case there exists an "*inverse*" mapping which associates with each element b of $\mathfrak{N}$ that element of $\mathfrak{M}$ which has b as its image.

Two sets that can be placed into a one-to-one correspondence are said to be (*cardinally*) *equivalent* or *equinumerous*, and we write:

$$\mathfrak{M} \sim \mathfrak{N}.$$

(Cardinally) equivalent sets are also said to have the same *cardinal number* or the same *cardinality*.

EXAMPLES. If to every positive integer n the number 0 or 1 is made to correspond, depending on whether n is even or odd, we have a mapping of the set of positive integers upon the set $\{0, 1\}$. This mapping, however, does not constitute a one-to-one correspondence. If every number n is mapped upon the number $2n$, then we have a one-to-one correspondence between the set of all positive integers and the set of all even integers. Thus the set of positive integers is (cardinally) equivalent to the set of all even integers.

EXERCISE. Prove the following three properties of the symbol $\sim$:

1. $\mathfrak{A} \sim \mathfrak{A}$.
2. $\mathfrak{A} \sim \mathfrak{B}$ implies $\mathfrak{B} \sim \mathfrak{A}$.
3. If $\mathfrak{A} \sim \mathfrak{B}$ and $\mathfrak{B} \sim \mathfrak{C}$, then $\mathfrak{A} \sim \mathfrak{C}$.

A proper subset may very well have the same cardinality as the set itself. This has been illustrated already by the second example above. Another example is the following: If we pair with every n the number $n-1$, then we have a one-to-one correspondence between the set of the positive integers and a set which contains also the zero besides the positive integers. In the following section we shall see, however, that the above mentioned possibility does not apply to "finite" sets.

3. THE NUMBER SEQUENCE

We presume that the reader is familiar with the set of natural numbers (positive integers)

$$1, 2, 3, \ldots$$

as well as with the following basic properties of this set (*Peano's axioms or postulates*)

I. 1 is a natural number.

II. Every natural number a has a definite successor (consequent) a^+ in the set of natural numbers.

III. We always have

$$a^+ \neq 1,$$

i.e. there is no number [3] with 1 as successor.

IV. If $a^+ = b^+$, then $a = b$,

i.e. for every number there exists no number, or exactly one number, with the former as successor.

V. "*The principle of complete induction.*" A set of natural numbers which contains the number 1 and which, for every number a it contains, contains its successor a^+ as well, contains all natural numbers.

The method of proof by *complete induction* rests upon property V. If it is to be proved that a property E is attributable to all numbers, it is first proved for the number 1, and then for an arbitrary n^+ under the "inducton hypothesis" that the property E is true for n. Then, by V., the set of numbers having the property E must contain all numbers.

SUM OF TWO NUMBERS. With every pair of numbers x, y there can be associated in exactly one way a natural number, called $x+y$, such that

(1) $x+1 = x^+$ for *every* x,

(2) $x+y^+ = (x+y)^+$ for every x and every y.[4]

This definition will permit us to write $a+1$ instead of a^+, henceforth. The following rules of arithmetic hold:

(3) $(a+b)+c = a+(b+c)$ ("Associative law of addition")

(4) $a+b = b+a$ ("Commutative law of addition")

(5) $a+b = a+c$ implies $b = c$.

PRODUCT OF TWO NUMBERS. With every pair of numbers x, y there can be associated in exactly one way a natural number, called $x \cdot y$ or xy, such that

(6) $x \cdot 1 = x$,

(7) $x \cdot y^+ = x \cdot y + x$ for every x and every y.

The following rules of arithmetic hold:

(8) $ab \cdot c = a \cdot bc$ ("Associative law of multiplication")

(9) $a \cdot b = b \cdot a$ ("Commutative law of multiplication")

(10) $a \cdot (b+c) = a \cdot b + a \cdot c$ ("Distributive law").

(11) $ab = ac$ implies $b = c$.

---

[3] For the present we shall frequently say "number" instead of "natural number" (positive integer).

[4] The proof of the above as well as the proofs of the following theorems may be found in E. Landau's little book *Grundlagen der Analysis*, Chapter 1. Leipzig, 1930.

GREATER AND LESS. If $a=b+u$, we may write $a>b$ or $b<a$. It is now shown that

(12) for any two numbers a, b one, and only one, of the relations

$$a<b, \quad a=b, \quad a>b.$$

holds.

(13) If $a<b$, and $b<c$, then $a<c$.

(14) If $a<b$, then $a+c<b+c$.

(15) If $a<b$, then $ac<bc$.

The solution u [unique according to (5)] of the equation $a=b+u$ for $a>b$ is designated by $a-b$. A short notation for "$a<b$ or $a=b$" is $a\leqq b$. Similarly, we write $a\geqq b$.

Furthermore, the following theorem holds:

Every non-empty set of natural numbers contains a least number, i.e. a number which is less than all other numbers of the set.

It is this theorem on which a second form of *complete induction* rests. If it is to be proved that a property E be true for all numbers, it need only be proved that it is true for an arbitrary number n, provided it is true for all numbers $<n$. (In particular, the property then holds for $n=1$, since there are no numbers <1 so that the "induction assumption" is superfluous.[5] (Of course, the inductive proof must also comprise the case $n=1$; otherwise it is insufficient.) Then the property E has to be true of all numbers; for otherwise the set of all numbers for which the property E does not hold would not be empty. Its smallest element would be a number n which would not have the property E, whereas all numbers $<n$ have the property E, giving a contradiction.

Besides the "proof by complete induction" in both its forms we have the "*definition* (or *construction*) *by complete induction.*" It is required to associate with every number x a new object $\varphi(x)$. A system of "recursion relations" is given, connecting every function value $\varphi(n)$ with the preceding values $\varphi(m)$ $(m<n)$. It is assumed that these relations determine every value $\varphi(n)$ uniquely, as soon as all the preceding $\varphi(m)$ $(m<n)$ are given and satisfy the given relations.[6] The simplest case is that in which for $m=n^+$ the value $\varphi(n^+)$ is expressed in terms of $\varphi(n)$ and where $\varphi(1)$ is given directly. Examples of this are the above relations (1),

[5] The assertion "all A's have the property B" is considered correct if there exist no A's at all. The assertion "E implies F" (where E and F are properties that may or may not be ascribed to certain objects x) will likewise be considered correct if there are no x's with the property E. All this is in conformity with our previous observation according to which the empty set is contained in every set.

Although this form of expression is not common usage, it is nevertheless expedient, insofar as only in this way is it possible to transmute the assertion "E implies F" into "not-E implies not-F." The negation of "E implies F" would be: There exists an x for which E holds and F does not.

[6] This assumption implies that $\varphi(1)$ is determined by the relations alone, since there are no numbers preceeding the 1.

(2), and (6), (7), respectively, which were used to define sums and products. Now the following proposition holds: *With given assumptions there is one, and only one, function* $\varphi(x)$ *whose values fulfill the given relations.*

PROOF: By a segment $(1, n)$ of the set of positive integers we shall mean the totality of the numbers $\leqq n$. First we show: In each segment $(1, n)$ there exists one, and only one, function $\varphi_n(x)$, which is defined for the numbers x of this segment, and which fulfills the given relations. This statement holds for the segment $(1, 1)$ as well as for every segment $(1, n^+)$, provided it holds for $(1, n)$; for, by virtue of the recursive formulae, the function value $\varphi(1)$ is uniquely determined, and so is the function value $\varphi(n^+)$ by the preceding values $\varphi(m) = \varphi_n(m)$ $(m \leqq n)$. Therefore, our proposition holds for every segment $(1, n)$. Thus we obtain a series of functions $\varphi_n(x)$. Each function $\varphi_n(x)$ is defined throughout $(1, n)$, hence for any smaller interval $(1, m)$, and since it fulfills the defining relations throughout the former segment, it will thus coincide with the function $\varphi_m(x)$. Therefore, two functions $\varphi_n(x)$, $\varphi_m(x)$ coincide for all values of x for which both are defined.

Now, the desired function $\varphi(x)$ must be defined on all segments $(1, n)$ and fulfill the defining relations, i.e. it must coincide with the function φ_n in all cases. Such a function φ exists and there is only one: its value $\varphi(x)$ is the common value of all $\varphi_n(x)$ defined for x. This proves the theorem.

We shall make frequent use of the "construction by complete induction."

EXERCISE. 1. Let a property E be true, first for $n = 3$, and second for $n+1$ whenever it is true for $n \geqq 3$. Prove that E is true for all numbers $\geqq 3$.

By incorporating the symbols $-a$ (negative integers) and 0 (zero) the sequence of positive integers can be extended to the domain of *whole numbers* (all *integers*). In order to explain the symbols $+$, $\cdot$, $<$ in this domain more conveniently, it is a matter of expediency to represent the whole numbers by pairs of natural numbers (a, b), namely

the natural number a by $(a+b, b)$,

the zero by (b, b),

the negative integer $-a$ by $(b, a+b)$,

where b is an arbitrary natural number in each case.

Every number may be represented by several symbols (a, b); but each symbol (a, b) defines one, and only one, whole number, viz.

the natural number $a-b$ if $a > b$,

the number 0 if $a = b$,

the negative integer $-(b-a)$ if $a < b$.

Now, we define:

$$(a, b) + (c, d) = (a+c, b+d),$$

$$(a, b) \cdot (c, d) = (ac+bd, ad+bc),$$

$(a, b) < (c, d)$ or $(c, d) > (a, b)$, if $a + d < b + c$,

and we verify easily: *first*, that the definitions are independent of the choice of the left-hand symbols, provided only that the numbers represented by these symbols remain the same; *second*, that the rules of arithmetic (3), (4), (5), (8), (9), (10), (12), (13), (14), and (15) are fulfilled for $c > 0$; *third*, that the solution of the equation $a + x = b$ in the extended domain is possible without restrictions and unique (the solution is again denoted by $b - a$); *fourth*, that $ab = 0$ holds when, and only when, either $a = 0$ or $b = 0$.[7]

EXERCISES. 2. Carry out the proofs of the above.

3. The same as Ex. 1 with the number 3 replaced by 0.

In the preceding sections only such elementary properties of the whole numbers have been mentioned as will play an important part in the subsequent chapters. The definition of fractions as well as the divisibility properties of the integers will be treated in Chapter III.

4. FINITE AND COUNTABLE (DENUMERABLE) SETS

A set having the same cardinality as the set of natural numbers is called *finite*. The empty set is also a finite set.[8]

This can be expressed more simply: A set is called finite if subscripts from 1 to n can be affixed to its elements so that distinct elements have distinct subscripts and that all subscripts from 1 to n are employed. Thus the elements of a finite set $\mathfrak{A}$ may be designated by $a_1, \ldots, a_n$: $\mathfrak{A} = \{a_1, \ldots, a_n\}$.

EXERCISE. 1. Prove by complete induction on n that any subset of a finite set $\mathfrak{A} = \{a_1, \ldots, a_n\}$ is itself finite.

A set which is not finite is called *infinite*. The set of all integers, for example, is an infinite set, as we shall presently prove.

The principal theorem on finite sets (also known as "Principal Theorem of Arithmetic") is expressed as follows:

A finite set cannot have the same cardinality as another set in which it is contained as a proper subset.

PROOF: Suppose a mapping of a finite proper subset $\mathfrak{A}$ upon a set $\mathfrak{D}$ containing $\mathfrak{A}$ as a proper subset were possible. Let the elements of the subset $\mathfrak{A}$ be $a_1, \ldots, a_n$. Let the images be $\varphi(a_1), \ldots, \varphi(a_n)$; among them will appear $a_1, \ldots, a_n$ and, moreover, at least one more element which we shall call a_{n+1}.

For $n = 1$ the absurdity is immediate: a single element a_1 cannot have two distinct images a_1, a_2.

[7] Reference is made to E. Landau's *Grundlagen der Analysis*, Chapter 4, where the negative numbers and the zero are introduced in a slightly different way.

[8] Other definitions of the concept of a finite set may be found in A. Tarski's "Sur les ensembles finis," *Fund. Math.* 6 (1925).

The impossibility of a mapping φ with the above properties being assumed for $n-1$, we shall now prove it for n.

We may assume that $\varphi(a_n) = a_{n+1}$ for if this is not so, e.g., if

$$\varphi(a_n) = a' \qquad (a' \neq a_{n+1})$$

then a_{n+1} has a different original a_i :

$$\varphi(a_i) = a_{n+1},$$

and instead of the mapping φ some other mapping may be devised, where the a_{n+1} is paired with the a_n , and the a' with the a_i , but where all other pairings are the same as in the mapping φ .

Now the function φ maps the subset $\mathfrak{A}' = \{a_1, \ldots, a_{n-1}\}$ upon the set $\varphi(\mathfrak{A}')$, which is obtained from $\varphi(\mathfrak{A}) = \mathfrak{D}$ by leaving out the element $\varphi(a_n) = a_{n+1}$.

Since $\varphi(\mathfrak{A}')$ thus contains $a_1, \ldots, a_n$, it contains $\mathfrak{A}'$ as a proper subset and is a unique image of $\mathfrak{A}'$. This result is impossible by the induction hypothesis.

An immediate consequence of this theorem is that a set can never be equivalent to two different segments of the number sequence; for otherwise the two segments would have to be equivalent to one another, which condition is impossible, since one is necessarily a proper subset of the other. Thus a finite set $\mathfrak{A}$ is equivalent to one, and only one, segment $(1, n)$ of the number sequence. The number n which is thus uniquely determined is called the *number of elements of the set* $\mathfrak{A}$ and can serve as a measure of cardinality.

Secondly, we infer that a segment of the number sequence cannot be equivalent to the entire number sequence. Thus the number sequence is infinite. Any set equivalent to the sequence of natural numbers is called *denumerably infinite.* Accordingly, the elements of a denumerably infinite set may be labeled with subscripts such that every natural number is used as a subscript exactly once.

Finite sets and denumerably infinite sets are both called *denumerable* or *countable.*

EXERCISES. 2. Prove that the number of elements of a union of two mutually exclusive finite sets is equal to the sum of the numbers of the individual sets. (Complete induction by means of the recursion formulae (1), (2), Section 3.)

3. Prove that the number of elements of a union of r sets any two of which are mutually exclusive and each of which contains s elements is equal to rs. (Complete induction by means of the recursion formulae (6), (7), Section 3.)

4. Prove that any subset of the number sequence is countable, and derive herefrom that a set is countable when, and only when, its elements can be labeled with subscripts such that distinct elements will have distinct subscripts.

AN EXAMPLE OF A NON-DENUMERABLE (UNCOUNTABLE) SET. The set of all denumerably infinite sequences of natural numbers is uncountable. It is readily seen that it is not finite. If it were denumerably infinite, then each sequence would have a subscript, and to each subscript i belongs a sequence which we may denote by

$$a_{i1}, a_{i2}, \ldots$$

Now let us write the following sequence

$$a_{11}+1, \qquad a_{22}+1, \ldots$$

This sequence would likewise have to have a subscript, say, the subscript j. Thus we would have

$$a_{j1}=a_{11}+1; \qquad a_{j2}=a_{22}+1, \text{ etc.};$$

in particular

$$a_{jj}=a_{jj}+1,$$

which is a contradiction.

EXERCISES. 5. Prove that the set of all integers (positive, negative, and zero) is denumerably infinite, and also that the set of the even numbers is denumerably infinite.

6. Prove that the set of all real numbers (i.e. of all non-terminating decimal fractions) is uncountable. (The method of proof is the same as in the above example.)

7. Prove that the cardinality of a denumerably infinite set does not change, if a finite number or a denumerably infinite number of new elements is added.

The union of a countable set of countable sets is itself countable.

PROOF: Let the countable sets be denoted by $\mathfrak{M}_1, \mathfrak{M}_2, \ldots$, and let the elements of $\mathfrak{M}_i$ be $m_{i1}, m_{i2}, \ldots$.

There are only a finite number of elements m_{ik} with $i+k=2$ and likewise a finite number of elements with $i+k=3$, etc. By first labeling the elements for which $i+k=2$ (say in the order of increasing values of i) and then (continuing in numerical order) those with $i+k=3$ etc., each element m_{ik} will eventually have a subscript, and distinct elements will have distinct subscripts. This proves the above assertion.

EXERCISE. 8. Prove that the set of all irreducible fractions $\pm\frac{a}{b}$ (a, b are relatively prime natural numbers) is denumerably infinite.

5. PARTITIONS

The equality sign satisfies the following conditions:

$$a=a.$$

$a=b$ implies $b=a$.

If $a=b$ and $b=c$, then $a=c$.

We may say instead: The relation $a=b$ is *reflexive, symmetric, and transitive.* If, among the elements of any set, a relation $a\sim b$ is defined (so that for each pair of elements a, b it is known whether or not $a\sim b$), and if this relation satisfies the same conditions,

1. $a\sim a$;
2. $a\sim b$ implies $b\sim a$;
3. if $a\sim b$ and $b\sim c$, then $a\sim c$,

then the relation $a \sim b$ is called an *equivalence relation*. For instance, the relation $\mathfrak{M} \sim \mathfrak{N}$ ($\mathfrak{M}$ cardinally equivalent to $\mathfrak{N}$), as defined in Section 2 for the sets $\mathfrak{M}, \mathfrak{N}, \ldots$ satisfies the conditions of these axioms. The congruence relation for triangles is also such a relation. A third example: Let two numbers in the domain of all integers be called equivalent if their difference is divisible by 2. Obviously, the conditions are fulfilled.

Now, whenever an equivalence relation is given, we may form a $\mathfrak{K}_a$ of all those elements which are equivalent to any element a. Then all elements of a class are equivalent to each other; for according to 2. and 3. it follows from $a \sim b$ and $a \sim c$ that $b \sim c$, and all elements equivalent to an element of the class are in the same class, since it follows from $a \sim b$ and $b \sim c$ that $a \sim c$. Thus a class is determined by any one of its elements: If, instead of from a, we start from any element b of the same class, we arrive at the same class: $\mathfrak{K}_b = \mathfrak{K}_a$. Accordingly, we may choose any b as a *representative* of the class.

However, if we proceed from an element b not belonging to the same partition (i.e. not equivalent to a), then $\mathfrak{K}_a$ and $\mathfrak{K}_b$ cannot have a common element; for from $c \sim a$ and $c \sim b$ it would follow that $a \sim b$, and hence $b \in \mathfrak{K}_a$. Thus the classes $\mathfrak{K}_a$ and $\mathfrak{K}_b$ are mutually exclusive in this case.

The classes cover the given set entirely, since every element a lies in a class, namely in $\mathfrak{K}_a$. Thus the set is divided into classes all mutually exclusive. This division is exemplified in our last example by the partition of the integers into even and odd numbers.

We have seen that $\mathfrak{K}_a = \mathfrak{K}_b$ when, and only when, $a \sim b$. By introducing classes instead of elements we may replace the equivalence relation $a \sim b$ by an equality relation $\mathfrak{K}_a = \mathfrak{K}_b$.

If, conversely, a given set $\mathfrak{M}$ is partitioned into mutually exclusive classes, we may define: $a \sim b$, if a and b belong to the same class. Obviously, the relation $a \sim b$ satisfies the conditions of the axioms 1, 2, 3.

CHAPTER II

GROUPS

CONTENTS: Explanation of fundamental concepts of the group theory which are essential for the entire book, such as group, subgroup, isomorphism, homomorphism, normal divisor, factor group.

6. THE CONCEPT OF A GROUP

DEFINITION: A non-empty set $\mathfrak{G}$ of any sort of elements (such as numbers, mappings, transformations) is said to be a *group* if the following four postulates are fulfilled:

1. A *rule of combination* is given which associates with every pair of elements a, b of $\mathfrak{G}$ a third element of the same set, which most frequently is called a *product* of a and b and which is denoted by ab or $a \cdot b$ (the product may depend on the order in which the factors are arranged; ab may or may not be equal to ba).

2. The associative law: If a, b, c are any elements of $\mathfrak{G}$, then

$$ab \cdot c = a \cdot bc.$$

3. There exists (at least) one element e in $\mathfrak{G}$, called the (left) *identity*, such that

$$ea = a \qquad \text{for every element } a \text{ of } \mathfrak{G}.$$

4. If a is an element of $\mathfrak{G}$, there exists (at least) one element a^{-1} in $\mathfrak{G}$, called the (left) *inverse* of a, such that

$$a^{-1}a = e.$$

A group is called *Abelian* if ab is always equal to ba (*commutative law*).

EXAMPLES. If the elements of the set are numbers, and if the rule of combination is ordinary multiplication, then the zero, for which there is no inverse, has to be excluded. All rational numbers $\neq 0$ form a group (the identity is the number 1); the numbers 1 and -1, or the number 1 all by itself form groups.

The group concept does not depend on the name of the group operation which may well be addition of numbers instead of multiplication, as long as the postulates 1. to 4. are fulfilled. One need only drop the name "product" for the element formed from a and b, and read "sum $a+b$" instead of "product $a \cdot b$." In this case the

identity will be the number 0, since $0+a=a$ for every a. Similarly, the inverse of a is $-a$, since $-a+a=0$. The associative law of addition,

$$a+(b+c)=(a+b)+c$$

is always fulfilled for numbers. Thus, a set of numbers is a group under the law of addition if, for every pair of numbers a and b in the set, it contains their sum also, the zero, and the inverse $-a$ for every number a. Such sets of numbers are also known as *number modules*. For example, all rational numbers form a module, or all integers, all even numbers, and finally the number 0 all by itself.

An example of a group whose elements are not numbers may be found in the totality of rotations of a plane or of a space about a fixed point. We compose two rotations A, B by performing them successively. If B is carried out first and then A, the same result (i.e. the same final position of all points of the space) may as well be obtained by a single rotation which is denoted by $A \cdot B$ or AB. A more precise algebraic definition of rotations and their components will be given in Volume II; here in this chapter the group of spatial rotations shall be mentioned only with a view to geometric intuition, as a first example of a group whose elements are not numbers. The group of rotations in space is at the same time a first example of a non-Abelian group; for it is by no means immaterial—as is geometrically evident—whether one performs first the rotation A and then B, or first B and then A. The fact that the associative law is fulfilled will later be exhibited as a special case of the associative law for all transformations. The identity of the rotation group is the identical transformation which leaves every point in its original position. The inverse of a given rotation is the rotation in the opposite sense, which cancels the former.

The rotation group is a special case of the more general conception of a *transformation group*. By a *transformation* or *permutation* of a set $\mathfrak{M}$ we shall understand a one-to-one reciprocal mapping of the set $\mathfrak{M}$ upon itself, i.e., an operation which associates with every element a of $\mathfrak{M}$ an image $s(a)$ such that every element of $\mathfrak{M}$ is the image of exactly one a. For $s(a)$ we may write sa as well. The elements of $\mathfrak{M}$ are the *objects* of the transformation s. The term transformation is most frequently used for infinite sets, the term permutation for finite sets.

If the set $\mathfrak{M}$ is finite, and if its elements are labeled with subscripts $1, 2, \ldots n$, then any permutation may be described completely by a symbol in which, below every subscript k, the subscript $s(k)$ of the image is written. For example, the symbol

$$s=\begin{pmatrix}1&2&3&4\\2&4&3&1\end{pmatrix}$$

is that permutation on the digits 1, 2, 3, 4 which carries 1 into 2, 2 into 4, 3 into 3, and 4 into 1.

If we first perform a transformation t and then [1] a transformation s on the

[1] The order of succession is a matter of agreement. Frequently, the order is inverted; st thus means: first s, then t. In this case it is a matter of expediency to write the transformation on the right hand side of the objects: as instead of $s(a)$.

image elements of t, we say that st is the *product* of the two transformations, i.e.,

$$st(a) = s(t(a)),$$

thus for $s = \begin{pmatrix} 1 2 3 4 \\ 2 4 3 1 \end{pmatrix}$, $t = \begin{pmatrix} 1 2 3 4 \\ 2 1 4 3 \end{pmatrix}$, the product $st = \begin{pmatrix} 1 2 3 4 \\ 4 2 1 3 \end{pmatrix}$. Similarly we have $ts = \begin{pmatrix} 1 2 3 4 \\ 1 3 4 2 \end{pmatrix}$.

The associative law:

$$(rs)t = r(st)$$

can be proved generally for transformations thus: Applying the transformations of both sides to an arbitrary object a, we obtain

$$(rs)t(a) = (rs)(t(a)) = r(s(t(a)))$$
$$r(st)(a) = r(st(a)) \quad = r(s(t(a))),$$

giving us the same result in either case.

The *identity* or *identical transformation* is that mapping I which maps each object upon itself:

$$I(a) = a.$$

Obviously, the identical transformation has the characteristic property of the identity of a group: $Is = s$ for every transformation s.

The *inverse transformation* of a transformation is that mapping which maps $s(a)$ into a, thus canceling the transformation s. Denoting the inverse transformation by s^{-1}, we have for every object a:

$$s^{-1}s(a) = a$$

and therefore

$$s^{-1}s = I.$$

It follows from what has just been proved that all postulates 1. to 4. are fulfilled for all permutations on a set $\mathfrak{M}$. Therefore all these permutations form a group. For a finite set $\mathfrak{M}$ with n elements the group of its permutations is called the *symmetric group* [2] $\mathfrak{S}_n$.

Moreover, it follows that every set $\mathfrak{G}$ of transformations of a set $\mathfrak{M}$ forms a group, provided merely that $\mathfrak{G}$ contains a) the product of each two transformations of $\mathfrak{G}$, b) the inverse for each transformation, and c) the identity. If the set is non-empty, then the postulate c) is superfluous; for if s is an arbitrary transformation in $\mathfrak{G}$, then, by b), s^{-1} belongs to the set $\mathfrak{G}$ and therefore, by a), $s^{-1}s = I$ does, too.

Let us now return to the general theory of groups.

For $ab \cdot c$ or $a \cdot bc$ we shall write briefly: abc.

From 3. and 4. follows

$$a^{-1}aa^{-1} = ea^{-1} = a^{-1};$$

---

[2] The reason for this name is the fact that the functions of $x_1, \ldots, x_n$, which remain invariant under all permutations of the group are the "symmetric functions."

if we multiply on the left by an inverse element of a^{-1} we have

$$eaa^{-1}=e$$

or

$$aa^{-1}=e;$$

thus, every left inverse is also a right inverse. At the same time we see that the inverse of a^{-1} is a again. Furthermore we infer:

$$ae=aa^{-1}a=ea=a;$$

thus, the left identity is also a right identity.

Now the *possibility of a right-and left-hand division* follows:

5. *The equation* $ax=b$ *as well as the equation* $ya=b$ *have solutions in* $\mathfrak{G}$ *where* a *and* b *are arbitrary elements of* $\mathfrak{G}$.

These solutions are $x=a^{-1}b$ and $y=ba^{-1}$, since

$$a(a^{-1}b)=(aa^{-1})b=eb=b,$$
$$(ba^{-1})a=b(a^{-1}a)=be=b$$

The *uniqueness of division* can be proved just as easily:

6. $ax=ax'$ and likewise $xa=x'a$ imply $x=x'$.

For $ax=ax'$ yields $x=x'$ when both members are multiplied by a^{-1} from the left. The second part of the above assertion may be proved in exactly the same manner.

In particular, we may infer the uniqueness of the identity (as the solution of the equation $xa=a$) and the uniqueness of the inverse (as the solution of the equation $xa=e$). The (only) identity is often denoted by 1.

The possibility of division (5.) is a postulate which has the power to replace the postulates 3. and 4. Let us, therefore, assume that 1., 2., and 5. hold, and let us first prove 3. We choose an element c, and by e we denote a solution of the equation $xc=c$. Then

$$ec=c.$$

For any a we now solve the equation

$$cx=a.$$

Then

$$ea=ecx=cx=a.$$

which proves 3. Postulate 4., on the other hand, follows directly from the solvability of $xa=e$.

Thus, we may always employ 1., 2., 5. as equivalent group postulates instead of 1., 2., 3., 4.

If $\mathfrak{G}$ *is a finite set, 5. may be replaced by 6.* Thus, we need not even assume the possibility of division, but merely its uniqueness (besides the postulates 1. and 2.)

PROOF: Let a be any element. With every element x we associate the element ax. According to 6., this pairing constitutes a one-to-one correspondence; i.e.

the set $\mathfrak{G}$ is mapped in a one-to-one manner on a subset, namely the set of all products ax. However, since $\mathfrak{G}$ is a finite set by hypothesis, it cannot be mapped on a proper subset in a one-to-one correspondence. Therefore, the totality of the elements ax has to be identical with $\mathfrak{G}$, i.e., each element b can be written in the form $b=ax$ as asserted in the first postulate 5. The solvability of $b=xa$ is proved in the same manner; hence 5. follows from 6.

The number of elements of a finite group is said to be the *order* of the group.

FURTHER RULES OF OPERATION. For the inverse of a product the following rule holds:

$$(ab)^{-1}=b^{-1}a^{-1}.$$

since

$$(b^{-1}a^{-1})ab=b^{-1}(a^{-1}ab)=b^{-1}b=e.$$

In dealing with Abelian groups the group operation is frequently written as an addition, i.e. we write $a+b$ instead of $a\cdot b$. The group is then called an *additive group* or a *module* (generalization of the above defined number modules). In this case the identity is designated by 0, since it is, like the zero in the domain of all integers, characterized by the property

$$0+a=a$$

Similarly, the inverse element of a in a module is denoted by $-a$.

A short notation for $a+(-b)$ is $a-b$, because this element is the solution of the equation $x+b=a$:

$$(a-b)+b=a+(-b+b)=a+0=a.$$

EXERCISES. 1. The Euclidean motions of space combined with reflections (i.e. those transformations that preserve all distances between pairs of points) form an infinite non-Abelian group.

2. Prove that the elements e, a form a group (Abelian) if the group operation is defined by

$$ee=e,\quad ea=a,\quad ae=a,\quad aa=e$$

NOTE. The law of combination of a group may be exhibited in a "multiplication table," a table with double entry, where for every two elements the product is listed. For example, the multiplication table for the group $\{e, a\}$ is:

	e	a
e	e	a
a	a	e

3. Construct a multiplication table for the group of all permutations on three digits.

COMPOSITE PRODUCTS (SUMS); POWERS. In the same fashion as we denoted $ab\cdot c$ briefly by abc, we shall define the *composite products* of several factors

$$\prod_{\nu=1}^{n}a_\nu=\prod_{1}^{n}a_\nu=a_1a_2\ldots a_n$$

If $a_1, \ldots, a_N$ are given, the recursive definition (for $n < N$) is

$$\begin{cases} \prod_1^1 a_\nu = a_1, \\ \prod_1^{n+1} a_\nu = \left(\prod_1^n a_\nu\right) \cdot a_{n+1}. \end{cases}^3$$

In particular, $\prod_1^3 a_\nu$ is our old $a_1 a_2 a_3$, and $\prod_1^4 = a_1 a_2 a_3 a_4 = (a_1 a_2 a_3) a_4$, etc.

We shall now prove the following rule solely by means of the associative law:

$$\prod_{\mu=1}^{m} a_\mu \cdot \prod_{\nu=1}^{n} a_{m+\nu} = \prod_{\nu=1}^{m+n} a_\nu \tag{1}$$

In words this may be expressed as follows: *The product of two composite products is equal to the composite product of all their factors in the same serial order.* For example,

$$(ab)(cd) = abcd$$

is a special case of (1).

For $n = 1$ formula (1) is clear (by definition of the Π symbol). Once it has been proved for a value n, we have for the next higher value $n+1$:

$$\begin{aligned} \prod_1^m a_\mu \cdot \prod_1^{n+1} a_{m+\nu} &= \prod_1^m a_\mu \left(\prod_1^n a_{m+\nu} \cdot a_{m+n+1}\right) \\ &= \left(\prod_1^m a_\mu \cdot \prod_1^n a_{m+\nu}\right) a_{m+n+1} \\ &= \left(\prod_1^{m+n} a_\mu\right) a_{m+n+1} = \prod_1^{m+n+1} a_\nu. \end{aligned}$$

This proves (1).

NOTE: $\prod_1^n a_{m+\nu}$ is also written in the form $\prod_{m+1}^{m+n} a_\nu$. Occasionally, if it is convenient, one may write $\prod_1^0 a_\nu = e$.

A product of n identical factors is called a *power*:

$$a^n = \prod_1^n a \quad \text{(in particular } a^1 = a,\ a^2 = aa, \text{ etc.)}.$$

From the theorem just proved it follows that

$$a^n \cdot a^m = a^{n+m}. \tag{2}$$

Furthermore:

$$(a^m)^n = a^{mn}. \tag{3}$$

We leave the proof (by complete induction) to the reader.

The rules (1), (2), (3), proved so far, required but the associative law for their proof and, therefore, they will be applied in the following sections to all kinds of

[3] The symbol ν denoting the variable index may of course be replaced by any other symbol without changing the meaning of the product.

domains in which products are defined and where the associative law holds (such as in the domain of the natural numbers), even though these domains may not be groups.

If the multiplication is communtative as well (Abelian groups), it can be proved that the value of a composite product is independent of the order in which the factors are arranged, or more precisely: *If φ is a one-to-one mapping of the segment $(1, n)$ of the natural numbers upon itself, then*

$$\prod_{\nu=1}^{n} a_{\varphi(\nu)} = \prod_{1}^{n} a_\nu .$$

PROOF: For $n=1$ the assertion is obvious; therefore, assume it to be correct for $n-1$. There exists a k which is mapped into n: $\varphi(k)=n$. Then

$$\prod_{1}^{n} a_{\varphi(\nu)} = \prod_{1}^{k-1} a_{\varphi(\nu)} \cdot a_{\varphi(k)} \cdot \prod_{1}^{n-k} a_{\varphi(k+\nu)} = \prod_{1}^{k-1} a_{\varphi(\nu)} \cdot \prod_{1}^{n-k} a_{\varphi(k+\nu)} \cdot a_{\varphi(k)} .$$ [4]

Defining a mapping ψ of the segment $(1, n-1)$ upon itself by

$$\psi(\nu) = \varphi(\nu) \qquad (\nu < k)$$
$$\psi(\nu) = \varphi(\nu+1) \qquad (\nu \geqq k)$$

we obtain

$$\prod_{1}^{n} a_{\varphi(\nu)} = \prod_{1}^{k-1} a_{\psi(\nu)} \prod_{1}^{n-k} a_{\psi(k-1+\nu)} \cdot a_n = \prod_{1}^{n-1} a_{\psi(\nu)} \cdot a_n ,$$

which, according to the induction hypothesis,

$$= \prod_{1}^{n-1} a_\nu \cdot a_n = \prod_{1}^{n} a_\nu .$$

From the rule just proved it follows that for Abelian groups a notation such as

$$\prod_{1 \leqq i < k \leqq n} a_{ik} ,$$

or

$$\prod_{i<k} a_{ik} \qquad (i = 1, \ldots, n;\ k = 1, \ldots, n),$$

is justified, which means that the set of the index pairs i, k with $1 \leqq i < k \leqq n$ shall be labeled in any arbitrary serial order and that then the product is formed.

In any group the 0-th and the negative powers of an element a may be defined as usual by

$$a^0 = 1 ,$$
$$a^{-n} = (a^{-1})^n ,$$

and we can prove without difficulty that the rules (2), (3) are valid for any integral exponents.

In an additive group we write of course $\sum_{1}^{n} a_\nu$ instead of $\prod_{1}^{n} a_\nu$, and accordingly,

[4] For $k=1$ the first factor drops out, for $k=n$ the second; this, however, has no effect on the proof.

$n \cdot a$ instead of a^n. In the additive group of the integers this definition is in agreement with that of the product of two integers. All proofs given for products may now be applied to sums.

Under the law of addition, rule (3) assumes the form of an associative law

$$n \cdot ma = nm \cdot a,$$

while (2) has the form of a "distributive law"

$$ma + na = (m+n)a.$$

Another distributive law may be added to the two preceding laws, viz.

$$m(a+b) = ma + mb$$

[for multiplication: $(ab)^m = a^m b^m$], which, however, holds only for Abelian groups. It may again be proved for m positive by induction.

For $m = 1$ the proof is clear. Assuming $m(a+b) = ma + mb$ we have

$$\begin{aligned}(m+1)(a+b) &= m(a+b) + (a+b)\\ &= ma + mb + a + b\\ &= (ma + a) + (mb + b)\\ &= (m+1)a + (m+1)b.\end{aligned}$$

As can be seen, the commutativity of mb and a is utilized for the proof. For $m = 0$ the assertion is likewise clear, while for negative m we need but apply the definition of negative powers in order to return to positive m.

EXERCISES. 5. Prove for Abelian groups that

$$\prod_{\nu=1}^{n} \prod_{\mu=1}^{m} a_{\mu\nu} = \prod_{\mu=1}^{m} \prod_{\nu=1}^{n} a_{\mu\nu}.$$

6. Similarly, prove

$$\prod_{\nu=1}^{n} \prod_{\mu=1}^{\nu} a_{\mu\nu} = \prod_{\mu=1}^{n} \prod_{\nu=\mu}^{n} a_{\mu\nu}.$$

7. The order of the symmetric group $\mathfrak{S}_n$ is $n! = \prod_{1}^{n} \nu$. [Complete induction on n].

7. SUBGROUPS

For a non-empty subset $\mathfrak{g}$ of a group $\mathfrak{G}$ to be itself a group, provided the rule of combination for the elements of $\mathfrak{g}$ is the same as that for the elements of $\mathfrak{G}$, it is necessary and sufficient that $\mathfrak{g}$ fulfill the postulates 1., 2., 3., and 4. Postulate 1. requires that, if a and b lie in $\mathfrak{g}$, ab has to lie in $\mathfrak{g}$, too. Postulate 2. is of course fulfilled for $\mathfrak{g}$, since it is fulfilled even for $\mathfrak{G}$. Postulates 3. and 4. state that the identity lies in $\mathfrak{g}$, and that, if $\mathfrak{g}$ contains a, it also contains the inverse element a^{-1}. Again, the requirement for the identity is superfluous; for if a is any element in $\mathfrak{g}$, then a^{-1} lies also in $\mathfrak{g}$, and so does the product $aa^{-1} = e$. Thus we have proved:

For a non-empty subset $\mathfrak{g}$ *of a given group* $\mathfrak{G}$ *to be a subgroup, the following conditions are necessary and sufficient:*

1. $\mathfrak{g}$ *contains the product* ab *for any two elements* a, b.

2. $\mathfrak{g}$ *contains the inverse* a^{-1} *to every element* a.

If, in particular, $\mathfrak{g}$ is finite, even the second of these requirements is unnecessary; for in this case 3. and 4. may be replaced by 6., and since postulate 6. applies to $\mathfrak{G}$, it will surely apply to $\mathfrak{g}$.

In general, postulates 1. and 2. can be replaced by one single postulate: If a and b are elements of $\mathfrak{g}$, $\mathfrak{g}$ shall also contain ab^{-1} for $\mathfrak{g}$ then contains $aa^{-1}=e$ with a, and furthermore, $ea^{-1}=a^{-1}$; hence with a and b, it contains also b^{-1} and $a(b^{-1})^{-1}=ab$.

If (in Abelian groups) the group relation is additive, a subgroup is characterized by the fact that it contains $a+b$, as soon as it contains a and b, and $-a$, as soon as it contains a. These two postulates are implied in the single posulate that with a and b the subgroup shall also contain $a-b$.

EXAMPLES OF SUBGROUPS:

Any group contains as a subgroup the identity group $\mathfrak{E}$ consisting only of the identity.

The most important subgroup of the symmetric group $\mathfrak{S}_n$ of all permutations on n objects is the *alternating group* $\mathfrak{A}_n$, containing these permutations which, when applied to the variables $x_1, \ldots, x_n$, carry the function

$$\Delta=\prod_{i<k}(x_i-x_k) \tag{1}$$

into itself. These permutations are called even, the others odd. Odd permutations reverse the sign of the function Δ. Any transposition (i.e. a permutation on two digits) is an odd permutation. The product of two even or of two odd permutations is even; the product of one even and one odd permutation is odd. It follows from the first property that $\mathfrak{A}_n$ is a group. Since a fixed transposition multiplied by an even permutation yields an odd one, and vice versa, there exist as many even permutations as there are odd ones; therefore, there are $\frac{n!}{2}$ of each kind (cf. Section 6, Ex. 7).

In order to facilitate the notation of the subgroups of the symmetric group $\mathfrak{S}_n$, the well-known symbols for cyclic or circular permutations are used:

By $(pqrs)$ we shall denote a cyclic permutation, carrying p into q, q into r, and r into s, and s into p, leaving the rest of the objects fixed. It is easy to show that any permutation is uniquely expressible (except for the order of succession) as a product of such cyclic permutations or "cycles" *as*

$$(ikl\ldots)\ (pq\ldots)\ldots,$$

where no two cycles have an element in common. The factors of this product commute. A cycle containing one element, say (1), is the identical permutation. It is understood that

$$(1\ 2\ 5\ 4)=(2\ 5\ 4\ 1),$$

etc.

Employing these symbols, we may represent the $3! = 6$ permutations of the group $\mathfrak{S}_3$ thus:

(1), (1 2), (1 3), (2 3), (1 2 3), (1 3 2)

It is easy enough to determine all the subgroups. They are (besides $\mathfrak{S}_3$ itself):

$$\left\{\begin{array}{l} \mathfrak{A}_3\colon (1),\ (1\ 2\ 3),\ (1\ 3\ 2); \\ \mathfrak{S}_2\colon (1),\ (1\ 2); \\ \mathfrak{S}_2'\colon (1),\ (1\ 3); \qquad \mathfrak{S}_2''\colon (1),\ (2\ 3); \\ \mathfrak{E}\colon (1). \end{array}\right.$$

If $a, b, \ldots$ are arbitrary elements of a group $\mathfrak{G}$, there may exist, besides $\mathfrak{G}$, other subgroups containing $a, b, \ldots$ The intersection of all these groups is another group $\mathfrak{A}$. It is called a group generated by $a, b, \ldots$. This group surely contains all such power products as $a^{-1}a^{-1}bab^{-1}\ldots$ (with a finite number of factors with or without repetition). But these power products form a group themselves which contains $a, b, \ldots$, and which thus comprehends $\mathfrak{A}$. Consequently, this group is identical with $\mathfrak{A}$. We have thus shown:

The group generated by a, b . . . consists of all power products, each of a finite number of these elements.

In particular, a single element a generates the group of all powers $a^{\pm n}$ (including $a^0 = e$). Since

$$a^n a^m = a^{n+m} = a^m a^n$$

this group is an Abelian group.

A group consisting of the powers of a single element is called *cyclic.*

Now there are two possibilities. Either all powers a^k are distinct; then the cyclic group

$$\ldots,\ a^{-2},\ a^{-1},\ a^0,\ a^1,\ a^2,\ \ldots$$

is *infinite.* Or it may occur that

$$a^h = a^k,\ h > k$$

Then

$$a^{h-k} = e \qquad (h-k>0).$$

In this case let n be the smallest positive exponent for which $a^n = e$. Then the powers $a^0, a^1, a^2, \ldots, a^{n-1}$ are all distinct from each other; for

$$a^h = a^k \qquad (0 \leqq k < h < n)$$

would imply

$$a^{h-k} = e \qquad (0 < h-k < n),$$

which is contrary to the assumption made as regards n.

If every integer m is represented in the form

$$m = qn + r \qquad (0 \leqq r < n)$$

then

$$a^m = a^{qn+r} = a^{qn}a^r = (a^n)^q a^r = e\,a^r = a^r.$$

Thus all powers of a are already contained in the sequence $a^0, a^1, \ldots, a^{n-1}$.

The cyclic group, therefore, has exactly n elements, viz.

$$a^0,\ a^1,\ \ldots,\ a^{n-1}.$$

The number n, which is the order of the cyclic group generated by a, is known as the *order of the element* a. If all powers of a are distinct from each other, a is called an element of *infinite order*.

EXAMPLES. The integers

$$\ldots, -2, -1, 0, 1, 2, \ldots$$

with addition as the rule of combination form an infinite cyclic group. The groups $\mathfrak{S}_2$, $\mathfrak{A}_3$ are cyclic groups of orders 2, 3.

EXERCISES. 1. In an Abelian group the product of an element a of order n and of an element b of order m is an element of order mn, provided m and n are numbers without a common divisor >1.

2. There are cyclic permutation groups of any given order.

3. Prove by induction on n that the $n-1$ transpositions $(1\,2), (1\,3), \ldots (1\,n)$ for $n>1$ generate the symmetric group $\mathfrak{S}_n$.

4. Prove, as in 3., that for $n>2$ the $n-2$ cyclic permutations on three digits $(1\,2\,3), (1\,2\,4), \ldots, (1\,2n)$ generate the alternating group $\mathfrak{A}_n$.

We shall now determine all subgroups of the cyclic groups. Let $\mathfrak{G}$ be a cyclic group generated by a, and let $\mathfrak{g}$ be a subgroup not containing the 1 alone. If $\mathfrak{g}$ contains an element a^{-m} with a negative exponent, it will also contain the inverse element a^m. Let a^m be the element of $\mathfrak{g}$ having the smallest positive exponent. We shall prove that all elements of $\mathfrak{g}$ are powers of a^m. If a^s is an arbitrary element of $\mathfrak{g}$, we have

$$s = qm + r \qquad (0 \leqq r < m)$$

Now $a^s(a^m)^{-q} = a^{s-mq} = a^r$ is an element of $\mathfrak{g}$, where $r<m$, whence it follows that $r=0$ because of the choice of m, and therefore $s=qm$ and $a^s=(a^m)^q$. Thus all elements of $\mathfrak{g}$ are powers of a^m.

If a is of finite order n so that $a^n=e$, then n must be divisible by m: $n=qm$, since $a^n=e$ is in the subgroup $\mathfrak{g}$. The subgroup $\mathfrak{g}$ then consists of the elements $a^m,\ a^{2m},\ \ldots,\ a^{qm}=e$ and is of order q. If, on the other hand, a is of infinite order, the subgroup $\mathfrak{g}$ consisting of the elements $e,\ a^{\pm m},\ a^{\pm 2m}, \ldots$ is likewise of infinite order. Thus we have proved the following:

A subgroup of a cyclic group is itself cyclic. It consists either of just the 1, *or of the powers of the element* a^m *with the smallest possible positive* m; *in other words, it consists of the* m*-th powers of the elements of the original group. For a cyclic group of infinite order,* m *may be chosen at will, whereas for a cyclic group of finite order* n *the number* m *must be a factor of* n. *In this case the subgroup is of order* $q=\frac{n}{m}$. *To every such number* m *belongs one, and only one, subgroup* $\{a^m\}$ *of the cyclic group* $\{a\}$.

8. COMPLEXES. COSETS

In group theory a *complex* is defined as an arbitrary set of elements of a group $\mathfrak{G}$.

By the *product* $\mathfrak{g}\mathfrak{h}$ of two complexes $\mathfrak{g}$ and $\mathfrak{h}$ we understand the set of all products gh where g is taken from $\mathfrak{g}$, and h from $\mathfrak{h}$. If in the product $\mathfrak{g}\mathfrak{h}$ one of the complexes, say $\mathfrak{g}$, consists of only one element g, we may simply write $g\mathfrak{h}$ instead of $\mathfrak{g}\mathfrak{h}$.

Obviously, the following rule holds:

$$\mathfrak{g}(\mathfrak{h}\mathfrak{k}) = (\mathfrak{g}\mathfrak{h})\mathfrak{k}.$$

In composite products of complexes the parentheses may therefore be omitted [cf. Section 6, (1)].

If the complex $\mathfrak{g}$ is a group, then

$$\mathfrak{g}\mathfrak{g} = \mathfrak{g}.$$

Let $\mathfrak{g}$ and $\mathfrak{h}$ be subgroups of $\mathfrak{G}$. The question arises: Under what conditions is the product $\mathfrak{g}\mathfrak{h}$ itself a group? The totality of the inverses of the elements of $\mathfrak{g}\mathfrak{h}$ is $\mathfrak{h}\mathfrak{g}$, since the inverse of gh is $h^{-1}g^{-1}$. Thus, for $\mathfrak{g}\mathfrak{h}$ to be a group, it is necessary that

$$\mathfrak{h}\mathfrak{g} = \mathfrak{g}\mathfrak{h} \tag{1}$$

i.e. $\mathfrak{g}$ must commute with $\mathfrak{h}$. This condition is also sufficient; for if it is satisfied, $\mathfrak{g}\mathfrak{h}$ contains for every gh also the inverse $h^{-1}g^{-1}$ and, moreover, for any two elements also their product, since

$$\mathfrak{g}\mathfrak{h}\mathfrak{g}\mathfrak{h} = \mathfrak{g}\mathfrak{g}\mathfrak{h}\mathfrak{h} = \mathfrak{g}\mathfrak{h}.$$

Restating the above we say: *The product* $\mathfrak{g}\mathfrak{h}$ *of two subgroups* $\mathfrak{g}$ *and* $\mathfrak{h}$ *of* $\mathfrak{G}$ *is itself a group when, and only when, the subgroups* $\mathfrak{g}$ *and* $\mathfrak{h}$ *are commutative.* It is of course not necessary that each element of $\mathfrak{g}$ be commutative with each element of $\mathfrak{h}$. If condition (1) is satisfied, then the product $\mathfrak{g}\mathfrak{h}$ is the group generated by $\mathfrak{g}$ and $\mathfrak{h}$.

In an Abelian group (1) is always satisfied. If we write the Abelian group under the law of addition, that is, if $\mathfrak{g}$ and $\mathfrak{h}$ are submodules of a module, we write $(\mathfrak{g}, \mathfrak{h})$ instead of $\mathfrak{g}\mathfrak{h}$, while we shall reserve the symbol $\mathfrak{g}+\mathfrak{h}$ for the special case of the "direct sum" to be investigated later on.

If $\mathfrak{g}$ is a subgroup, and a an element of $\mathfrak{G}$, then the complex $a\mathfrak{g}$ is called *a left coset*, and the complex $\mathfrak{g}a$ *a right coset* (or a *residue class*) of $\mathfrak{g}$ in $\mathfrak{G}$.

If a lies in $\mathfrak{g}$, we have $a\mathfrak{g} = \mathfrak{g}$; hence one of the left (and also one of the right) cosets of $\mathfrak{g}$ is always equal to $\mathfrak{g}$ itself.

In the following we shall mainly be concerned with left cosets, although all our considerations are valid for right cosets as well.

Two cosets $a\mathfrak{g}$, $b\mathfrak{g}$ can very well be equal without a being equal to b; for whenever $a^{-1}b$ lies in $\mathfrak{g}$, we have

$$b\mathfrak{g} = aa^{-1}b\mathfrak{g} = a(a^{-1}b\mathfrak{g}) = a\mathfrak{g}.$$

Two *different* cosets have no common element; for if the cosets $a\mathfrak{g}$ and $b\mathfrak{g}$ have a common element, say

$$ag_1 = bg_2,$$

it follows that

$$g_1 g_2^{-1} = a^{-1}b,$$

so that $a^{-1}b$ lies in $\mathfrak{g}$; accordingly, $a\mathfrak{g}$ and $b\mathfrak{g}$ are identical.

Every element a belongs to a coset, namely to the coset $a\mathfrak{g}$. The latter surely contains the element $ae = a$. According to what has just been proved, the element a belongs to *just one* coset. Therefore, we may regard any element a as a *representative* of the coset $a\mathfrak{g}$ containing a.

According to the preceding paragraphs the cosets constitute a partition of the group $\mathfrak{G}$. Every element belongs to one, and only one, class.[5]

Any two cosets are equipotent; for $ag \to bg$ defines a one-to-one mapping of $a\mathfrak{g}$ upon $b\mathfrak{g}$.

The cosets, except $\mathfrak{g}$ itself, do not constitute groups, since they do not contain the identity.

The number of the various cosets of a subgroup $\mathfrak{g}$ in $\mathfrak{G}$ is called the *index* of $\mathfrak{g}$ in $\mathfrak{G}$. The index can be finite or infinite.

If N is the order of $\mathfrak{G}$ (assumed finite), n the order of $\mathfrak{g}$, and j the index, the following relation holds:

$$N = jn; \tag{2}$$

for $\mathfrak{G}$ is divided into j partitions each of which contains n elements.[6]

For finite groups the index j may be computed from (2):

$$j = \frac{N}{n}.$$

Theorem: The order of a finite group is divisible by the order of each one of its subgroups.[7]

If, in particular, we consider a cyclic group generated by an element c as the subgroup, we have the following corollary:

The order of an element of a finite group is a factor of the order of the group.

An immediate consequence of this theorem is that *in a group of n elements the relation $a^n = e$ holds for every a.*

---

[5] Galois' notation

$$\mathfrak{G} = a_1\mathfrak{g} + a_2\mathfrak{g} + \cdots$$

is often found in literature, which means that the partitions are mutually exclusive and form the group $\mathfrak{G}$. We avoid this notation because we want to reserve the $+$ symbol for the direct sum to be discussed later on.

[6] It is true that the relation also holds when N is infinite; in this case, however, in order to explain its meaning, one has to introduce cardinal numbers, which we have not done.

[7] Translator's note: This theorem is also known as Lagrange's Theorem.

It may happen that all left cosets $a\mathfrak{g}$ are likewise right cosets. For this condition to be the case, it is necessary that the left coset containing an arbitrarily given element a be identical with the right coset containing a; i.e., for every a we must have

$$a\mathfrak{g} = \mathfrak{g}a \tag{3}$$

A subgroup $\mathfrak{g}$ having the property (3), i.e. commuting with every element a of $\mathfrak{G}$, is called *a normal*[8] *divisor, or a self-conjugate or invariant subgroup in* $\mathfrak{G}$.

If $\mathfrak{g}$ is a normal divisor, then the product of two cosets is itself a coset:

$$a\mathfrak{g} \cdot b\mathfrak{g} = a \cdot \mathfrak{g}b \cdot \mathfrak{g} = ab\mathfrak{g}\mathfrak{g} = ab\mathfrak{g}.$$

EXERCISES. 1. Find the right and left cosets for the subgroups of the $\mathfrak{S}_3$ group. Which of these subgroups are normal divisors?

2. Show that for any subgroup the inverses of the elements of a left coset form a right coset. Conclude from this that the index may also be determined as the number of the right cosets.

3. Show that any subgroup of index 2 is a normal divisor. Example: The alternating group of the symmetric group of n letters.

4. A subgroup of an Abelian group is always a normal divisor.

5. Those elements of a group which commute with all the elements of the group constitute a normal subgroup (the "center" or "central" of the group).

6. If $\mathfrak{G}$ is a cyclic group generated by a, and $\mathfrak{g}$ a subgroup distinct from $\mathfrak{E}$ and generated by a^m with the smallest m (see Section 7), then $1, a, a^2, \ldots, a^{m-1}$ are representatives of the cosets, and m is the index of $\mathfrak{g}$ in $\mathfrak{G}$.

7. If the product of any two left cosets of $\mathfrak{g}$ in $\mathfrak{G}$ is itself a left coset, $\mathfrak{g}$ is a normal subgroup in $\mathfrak{G}$.

9. ISOMORPHISMS AND AUTOMORPHISMS

Let two sets $\mathfrak{M}$ and $\overline{\mathfrak{M}}$ be given. Let there be defined any sort of relations between the elements of each of these sets. We may e.g. imagine that the sets $\mathfrak{M}$ an $\overline{\mathfrak{M}}$ are groups, and that the relations are the equations $a \cdot b = c$, which exist by virtue of the group property; or we may think of ordered sets and that the relations are $a > b$.

If it is possible to place the two sets into one-to-one correspondence such that the mapping preserves the relations, i.e., if with every element a of $\mathfrak{M}$ there can be associated an element $\bar{a}$ of $\overline{\mathfrak{M}}$ in a biunique manner so that the relations existing between any elements $a, b, \ldots$ of $\mathfrak{M}$ also exist between the associated elements $\bar{a}, \bar{b}, \ldots$ and vice versa, then the two sets are called *isomorphic* (with respect to the

[8] Here "divisor" means subgroup. "Normal" expresses the special property $a\mathfrak{g} = \mathfrak{g}a$. Translator's note: This group is sometimes called a "distinguished" subgroup, a literal translation of the German "ausgezeichnet."

relations in question), and we write $\mathfrak{M} \cong \overline{\mathfrak{M}}$. The mapping itself is called *isomorphism.*

In order to emphasize the one-to-one correspondence, we also say 1-*isomorphic* and 1-*isomorphism.*

Thus we can speak of 1-*isomorphic groups,* of isomorphically ordered or *similarly ordered sets,* etc. A 1-isomorphism of two groups is therefore a one-to-one mapping $a \to \bar{a}$, where $ab = c$ implies $\bar{a}\,\bar{b} = \bar{c}$ (and vice versa), and where therefore the product ab is always associated with the product $\bar{a}\,\bar{b}$.

Just as, in the general theory of sets, sets of the same cardinality are equivalent, so, in the theory of order types, similar sets, and in group theory, isomorphic groups are not to be regarded as substantially different from one another. Concepts and theorems which can be defined and proved on the basis of the given relations of a set may be applied directly to any 1-isomorphic set. For example, a set for which product relations are defined and which is 1-isomorphic with a group is itself a group, and the 1-isomorphism makes the identity, the inverse and subgroups reappear as identity, inverse and subgroups.

If, in particular, the two sets $\overline{\mathfrak{M}}$ and $\mathfrak{M}$ are identical, i.e., if the mapping under consideration associates with every element a an element $\bar{a}$ of the same set in a one-to-one manner, and if the relations are preserved, then the mapping is called *automorphism* or 1-*automorphism*).

If, for example, the set of integers is considered an ordered set, the mapping

$$a \to a + 1$$

is a 1-automorphism, since it maps the set of the integers biuniquely upon itself, and $a < b$ implies $a + 1 < b + 1$, and vice versa.

The 1-automorphisms of a set are an expression of its symmetry; for what is meant by symmetry, such as the symmetry of a geometric figure? It means that, under certain transformations (reflections, rotations, etc.), the figure is mapped upon itself, whereby certain relations (such as distances, angles, relative locations) are preserved; or, if we use our own terminology, we may say that the figure permits certain 1-automorphisms relative to its metric properties.

Obviously, the product of two 1-automorphisms (product of transformations according to Section 6) is itself a 1-automorphism, and the inverse operation of an automorphism is again a 1-automorphism. According to Section 6, it follows that the 1-automorphisms of an arbitrary set (with arbitrary relations between its elements) form a transformation group, the *automorphism group* of the set.

In particular, the 1-automorphisms of a group form a group themselves. Let us investigate some of these automorphisms a little more closely.

Let a be a fixed group element. The mapping carrying x into

$$\bar{x} = a x a^{-1} \tag{1}$$

is a 1-automorphism, since firstly (1) may be solved uniquely for x:

$$x = a^{-1} \bar{x}\, a;$$

which implies that the mapping is one-to-one; secondly, we have

$$\overline{x}\,\overline{y} = axa^{-1} \cdot aya^{-1} = a(xy)a^{-1} = \overline{xy};$$

therefore, the mapping is isomorphic.

The elements x, axa^{-1} are called *conjugate group elements.* The automorphisms $x \to axa^{-1}$ generated by the elements a are called *inner automorphisms* of the group. All other automorphisms (if they exist) are called *outer automorphisms.*

An inner automorphism $x \to axa^{-1}$ transforms a subgroup $\mathfrak{g}$ into a subgroup $a\mathfrak{g}a^{-1}$, which is said to be a *conjugate of* $\mathfrak{g}$ or *conjugate to* $\mathfrak{g}$.

If a subgroup $\mathfrak{g}$ is identical with all its conjugate subgroups, i.e.,

$$a\mathfrak{g}a^{-1} = \mathfrak{g} \qquad \text{for } \textit{every } a, \tag{2}$$

it simply means that the group $\mathfrak{g}$ commutes with every element a, i.e.,

$$a\mathfrak{g} = \mathfrak{g}a$$

and is, therefore, a *normal divisor* (Section 8). Restating the above, we may say:

The subgroups invariant under all inner automorphisms are normal divisors.

This theorem explains why normal divisors are also called "invariant subgroups" or "self-conjugate subgroups."

Postulate (2) may be replaced by the slightly weaker postulate

$$a\mathfrak{g}a^{-1} \leqq \mathfrak{g}; \tag{3}$$

for if (3) is true for every a, it will be true for a^{-1} as well:

$$a^{-1}\mathfrak{g}a \leqq \mathfrak{g},$$
$$\mathfrak{g} \leqq a\mathfrak{g}a^{-1};$$

now from (3) and (4) follows (2). Thus we may say:

A subgroup is a normal divisor if it contains with any element b all its conjugate element aba^{-1} as well.

EXERCISES. 1. Abelian groups have no inner automorphisms except the identity automorphism. Show that the group e, a, b, c with e as the identity and the rules of combination,

$$\begin{cases} a^2 = b^2 = c^2 = e, \\ ab = ba = c, \\ bc = cb = a, \\ ca = ac = b \end{cases}$$

has no inner automorphism except the identity automorphism, but that it has five outer automorphisms.

2. In permutation groups the transform aba^{-1} of an element b can be obtained by expressing b as a product of cycles (Section 7), and by performing the permutation a on the digits of these cycles. Give the proof! Use this proposition to compute aba^{-1} for

$$b = (1\,2)\,(3\,4\,5),$$
$$a = (2\,3\,4\,5).$$

3. Prove that the symmetric group $\mathfrak{S}_3$ has no outer, but six inner automorphisms.

4. The symmetric group $\mathfrak{S}_4$ has, besides itself and the identity group, only the following normal divisors:

a) the alternating group $\mathfrak{A}_4$,

b) Klein's four-group[9] $\mathfrak{B}_4$, consisting of the permutations

$$(1),\ (1\ 2)(3\ 4),\ (1\ 3)(2\ 4),\ (1\ 4)(2\ 3)$$

The latter group is Abelian and isomorphic with the group abstractly defined in Ex. 1.

5. If $\mathfrak{g}$ is a normal divisor in $\mathfrak{G}$, and if $\mathfrak{H}$ is an intermediate group

$$\mathfrak{g} \subseteq \mathfrak{H} \subseteq \mathfrak{G}$$

then $\mathfrak{g}$ is likewise a normal divisor in $\mathfrak{H}$.

6. All infinite cyclic groups are isomorphic with the additive group of integers.

7. The elements x of a group $\mathfrak{G}$ commuting with an element a, i.e., the elements x for which

$$xa = ax$$

form a group known as the *normalizer* of a. This group contains as a normal subgroup the cyclic group generated by a. The number of the elements conjugate to a is equal to the index of the *normalizer* in $\mathfrak{G}$.

8. The elements of a group $\mathfrak{G}$ may be divided into classes of conjugate elements. The number of the elements of a class is a factor of the order of $\mathfrak{G}$ if $\mathfrak{G}$ is finite. The identity forms a class by itself, and so does every element of the central (Section 8, Ex. 5).

9. If in a group of order p^n, where p is a prime number,[10] a_λ is the number of the classes with p^λ elements, and, in particular, a_0 the number of the elements of the central, then

$$p^n = a_0 + a_1 p + a_2 p^2 + \dots$$

Show by means of this equation that the central of a group of order p^n cannot consist of the identity alone.

10. HOMOMORPHISM. NORMAL DIVISORS. FACTOR GROUPS

If, in two sets $\mathfrak{M}$ and $\overline{\mathfrak{M}}$ certain relations are defined, and if with every element a of $\mathfrak{M}$ there is associated exactly one image $\bar{a}$ in $\overline{\mathfrak{M}}$ such that

1) every element $\bar{a}$ of $\overline{\mathfrak{M}}$ appears as an image at least once,

2) all relations between the elements of $\mathfrak{M}$ hold for the corresponding elements

[9] Translator's note: This group is also known as fours-group, axial group, quadratic group, Vierergruppe, etc.

[10] Translator's note: This group is called a prime power group.

of $\overline{\mathfrak{M}}$, then this mapping is called a *homomorphism*, and $\overline{\mathfrak{M}}$ is said to be *homomorphic with* $\mathfrak{M}$.

In symbols, this is expressed as $\mathfrak{M} \sim \overline{\mathfrak{M}}$. If $\overline{\mathfrak{M}} \leqq \mathfrak{M}$, i.e. if the homomorphism is a mapping of $\mathfrak{M}$ into itself, it is called an *endomorphism*.

If the mapping is biunique, and if the property of homomorphism holds in the opposite direction as well, we again have a 1-isomorphism, as defined above.

Under a homomorphic mapping the elements of $\mathfrak{M}$ that have a fixed image $\bar{a}$ in $\overline{\mathfrak{M}}$ may be united in a class $\mathfrak{a}$. Every element a belongs to one, and only one, class $\mathfrak{a}$, i.e., the set $\mathfrak{M}$ is *divided into classes* that have a one-to-one correspondence with the elements of $\overline{\mathfrak{M}}$.

EXAMPLES. If every element of a group is mapped upon the identity, a homomorphism of the group with the identity group is produced. Homomorphism can also be obtained by associating with each permutation of a permutation group the number $+1$ or -1, depending on whether the permutation is even or odd; the associated group is the multiplicative group of the numbers $+1$ and -1.

If we associate with every integer m the power a^m of an element a of a group, we obtain a homomorphism of the additive group of integers with the cyclic group generated by a, since the product $a^{m+n} = a^m \cdot a^n$ is associated with the sum $m+n$. If a is an element of infinite order, the homomorphism is an isomorphism.

Now we shall, in particular, investigate homomorphisms of groups.

If, in a set $\overline{\mathfrak{G}}$, *products* $\bar{a}\,\bar{b}$ *(i.e., relations of the form* $\bar{a}\,\bar{b} = \bar{c}$*) are defined, and if a group* $\mathfrak{G}$ *is mapped homomorphically upon* $\overline{\mathfrak{G}}$, *then* $\overline{\mathfrak{G}}$ *is itself a group; or briefly: the homomorphic image of a group is itself a group.*

PROOF: First of all, any three given elements $\bar{a}$, $\bar{b}$, $\bar{c}$ of $\overline{\mathfrak{G}}$ are always images of elements of $\mathfrak{G}$, say of a, b, c. From

$$ab \cdot c = a \cdot bc$$

now follows

$$\bar{a}\,\bar{b} \cdot \bar{c} = \bar{a} \cdot \bar{b}\,\bar{c}.$$

Furthermore,

$$ae = a \quad \text{for all } a$$

implies

$$\bar{a}\,\bar{e} = \bar{a} \quad \text{for all } \bar{a},$$

and

$$ba = e \qquad (b = a^{-1})$$

implies

$$\bar{b}\,\bar{a} = \bar{e}.$$

Thus there is an identity $\bar{e}$ in $\overline{\mathfrak{G}}$ and an inverse for every $\bar{a}$. Therefore $\overline{\mathfrak{G}}$ is a group. At the same time we have proved the following:

Under any homomorphism the identity goes into the identity, and inverses go into inverses.

We shall now study in greater detail the partition effected by a homomorphic

mapping $\mathfrak{G} \to \overline{\mathfrak{G}}$, and we shall find a very important one-to-one relation between homomorphisms and normal divisors.

The class $\mathfrak{e}$ *of* $\mathfrak{G}$, *to which corresponds the identity* $\bar{e}$ *of* $\overline{\mathfrak{G}}$ *under the homomorphism* $\mathfrak{G} \sim \overline{\mathfrak{G}}$, *is a normal divisor of* $\mathfrak{G}$, *and the class corresponding to any element of* $\overline{\mathfrak{G}}$ *is always a coset.*

PROOF: In the first place $\mathfrak{e}$ is a group; for if the homomorphism carries both a and b into $\bar{e}$, then it carries ab into $\bar{e}^2 = \bar{e}$; thus $\mathfrak{e}$ contains with any two elements their product. Furthermore, a^{-1} is mapped upon $\bar{e}^{-1} = \bar{e}$; thus $\mathfrak{e}$ contains also the inverse of every element.

The elements of a left coset $a\mathfrak{e}$ are all carried into the element $\bar{a}\,\bar{e} = \bar{a}$. If, conversely, an element a' is mapped into $\bar{a}$, then let x be determined from

$$a x = a'.$$

It follows that

$$\bar{a}\,\bar{x} = \bar{a},$$

$$\bar{x} = \bar{e}.$$

Hence x lies in $\mathfrak{e}$, and a' lies in $a\mathfrak{e}$.

The class of $\mathfrak{G}$, corresponding to the element $\bar{a}$, is thus seen to be exactly the left coset $a\mathfrak{e}$.

We may show in like manner that the partition corresponding to $\bar{a}$ must be the right coset $\mathfrak{e}a$. Therefore, right and left cosets coincide:

$$a\mathfrak{e} = \mathfrak{e}a,$$

and $\mathfrak{e}$ is a normal divisor. This completes the proof.

We started with a given homomorphism and arrived at a normal divisor. Let us now reverse the question: *Let a normal divisor* $\mathfrak{g}$ *of* $\mathfrak{G}$ *be given. Is it possible to form a group* $\overline{\mathfrak{G}}$ *homomorphic with* $\mathfrak{G}$ *so that the cosets of* $\mathfrak{g}$ *correspond exactly to the elements of* $\overline{\mathfrak{G}}$ *?*

In order to achieve this, we simply choose the cosets of $\mathfrak{g}$ themselves as elements of the group $\overline{\mathfrak{G}}$ to be constructed. According to Section 8, the product of two cosets of the normal divisor $\mathfrak{g}$ is itself a coset, and if a belongs to the coset $a\mathfrak{g}$ and b to $b\mathfrak{g}$, then ab belongs to $ab\mathfrak{g} = a\mathfrak{g} \cdot b\mathfrak{g}$, the coset of the product. It follows that the cosets form a set homomorphic with $\mathfrak{G}$ *and hence a group homomorphic with* $\mathfrak{G}$. This group is called a *quotient* of $\mathfrak{G}$ *by* $\mathfrak{g}$ or a *factor group* of $\mathfrak{G}$ or a *quotient group* of $\mathfrak{G}$. It is denoted by

$$\mathfrak{G}/\mathfrak{g}$$

The order of $\mathfrak{G}/\mathfrak{g}$ is equal to the index of $\mathfrak{g}$.

Here we realize the fundamental importance of normal divisors: they enable us to construct new groups homomorphic with given groups.

In the case of a homomorphic mapping of a group $\mathfrak{G}$ upon another group $\overline{\mathfrak{G}}$ we saw already that the cosets of a normal divisor $\mathfrak{e}$ in $\mathfrak{G}$ correspond (biuniquely) to the elements of $\overline{\mathfrak{G}}$. This correspondence is of course an isomor-

phism; for if $a\mathfrak{g}$, $b\mathfrak{g}$ are two cosets, $ab\mathfrak{g}$ is their product; the corresponding elements in $\overline{\mathfrak{G}}$ are $\bar{a}$, $\bar{b}$, $\overline{(ab)}$, and as a matter of fact

$$\overline{(ab)} = \bar{a} \cdot \bar{b}$$

because of the homomorphism. Thus we have

$$\mathfrak{G}/\mathfrak{e} \simeq \overline{\mathfrak{G}},$$

and the law of homomorphism for groups:

Any group $\overline{\mathfrak{G}}$ that is a homomorphic image of $\mathfrak{G}$ is isomorphic with a factor group $\mathfrak{G}/\mathfrak{e}$, where $\mathfrak{e}$ is the normal divisor of $\mathfrak{G}$ corresponding to the identity in $\overline{\mathfrak{G}}$. Conversely, $\mathfrak{G}$ is mapped homomorphically upon every factor group $\mathfrak{G}/\mathfrak{e}$, where $\mathfrak{e}$ is a normal divisor.

EXERCISES. 1. Trivial factor groups of any group $\mathfrak{G}$ are:

$$\mathfrak{G}/\mathfrak{E} \simeq \mathfrak{G}; \quad \mathfrak{G}/\mathfrak{G} \simeq \mathfrak{E}.$$

2. The factor group of the alternating group $(\mathfrak{S}_n/\mathfrak{A}_n)$ is a cyclic group of order 2.

3. The factor group $\mathfrak{S}_4/\mathfrak{B}_4$ of the four-group (Section 9, Ex. 4) is isomorphic with $\mathfrak{S}_3$.

4. The elements $aba^{-1}b^{-1}$ of a group $\mathfrak{G}$ and their products form a group called the *commutator subgroup* of $\mathfrak{G}$. It is a normal divisor, and its factor group is an Abelian group. Any normal divisor whose factor group is Abelian comprehends the commutator subgroup.

5. The elements commuting with a subgroup $\mathfrak{g}$ form a group $\mathfrak{h}$ containing $\mathfrak{g}$ as a normal divisor. This group is called the *normalizer* of $\mathfrak{g}$. The index of $\mathfrak{h}$ in $\mathfrak{G}$ is equal to the number of the subgroups of $\mathfrak{G}$ conjugate to $\mathfrak{g}$.

6. If $\mathfrak{G}$ is cyclic, a the generating element of $\mathfrak{G}$, and $\mathfrak{g}$ a subgroup of index m, then $\mathfrak{G}/\mathfrak{g}$ is cyclic of order m. The elements $1, a, a^2, \ldots, a^{m-1}$ may be chosen as the representatives of the cosets.

In an Abelian group every subgroup is a normal subgroup (cf. Section 8, Ex. 4). If the law of combination is addition, the groups and their subgroups are known as *modules*, as has already been said. The cosets $a + \mathfrak{M}$ (where $\mathfrak{M}$ is a module) are called residue classes modulo $\mathfrak{M}$, and the factor group $\mathfrak{G}/\mathfrak{M}$ is called a *residue class module of* $\mathfrak{G}$ modulo $\mathfrak{M}$.

Two elements a, b lie in a residue class if their difference lies in $\mathfrak{M}$. Two such elements are said to be *congruent modulo* $\mathfrak{M}$, and we write

$$a \equiv b \pmod{\mathfrak{M}}$$

or briefly

$$a \equiv b \ (\mathfrak{M}).$$

If a and b are congruent (modulo $\mathfrak{M}$), the residue classes $\bar{a}$ and $\bar{b}$ are identical. Conversely, $\bar{a} = \bar{b}$ always implies $a \equiv b(\mathfrak{M})$.

For example, the multiples of a positive integer m form a module in the domain of all integers, and we write accordingly

$$a \equiv b\,(m),$$

if the difference $a - b$ is divisible by m. The residue classes may be represented by $0, 1, 2, \ldots, m-1$, and the residue class module is a cyclic additive group of order m.

EXERCISE. 7. Any cyclic group of order m is isomorphic with the residue class module modulo an integer m.

CHAPTER III

RINGS AND FIELDS

CONTENTS: Definition of the concepts of a ring, an integral domain, a field. General methods for forming rings (or fields, respectively) from other rings. Theorems on factorization into primes in integral domains.

The concepts of this chapter will be used throughout the book.

11. RINGS

The quantities employed in algebraic and arithmetic operations vary in nature; at times we use the integers, or the rational, the real, complex or algebraic numbers, and at other times we deal with polynomials, or rational functions in n variables, etc. Later on we shall become familiar with quantities of a completely different nature, such as hypercomplex numbers, residue classes, etc., with which we can operate in the same or almost the same manner as with numbers. It is, therefore, desirable to arrive at a common concept embracing all these domains, and to investigate the rules of operation in these domains in general.

By a *system of double composition* we shall mean a set of elements in which, for any two elements $a, b, \ldots$, a sum $a+b$ and a product $a \cdot b$ belonging to the set are uniquely defined.

A system of double composition will be called a *ring* if the following *rules of operation* are satisfied for all elements of the system:

I. *Laws of addition.*

a) *Associative law:* $a+(b+c)=(a+b)+c$.

b) *Commutative law:* $a+b=b+a$.

c) *Solvability*[1] *of the equation* $a+x=b$ for all a and b.

II. *Law of multiplication.*

a) *Associative law:* $a \cdot bc = ab \cdot c$.

III. *Distributive laws.*

a) $a \cdot (b+c) = ab+ac$.

b) $(b+c) \cdot a = ba+ca$.

[1] A unique solution is not required.

ADDENDUM: A ring is called *commutative* if the commutative law holds for the multiplication:

II. b) $a \cdot b = b \cdot a$,

For the present we shall mainly deal with commutative rings.

REGARDING THE LAWS OF ADDITION. The three laws Ia, b, c merely assert that the ring elements form an Abelian group under addition.[2] Consequently, we can apply to rings all theorems proved for Abelian groups. There exists one (and only one) *null element* 0 such that

$$a + 0 = a \quad \text{for every } a.$$

Moreover, for every element a, there exists an *inverse element* $-a$ such that

$$-a + a = 0.$$

Furthermore, the equation $a + x = b$ is not only solvable, but the solution is unique; its sole solution is

$$x = -a + b;$$

which we denote also by $b - a$. By virtue of

$$a - b = a + (-b)$$

any difference may be written as a sum. Hence the same laws of associativity hold for differences as well as for sums, e.g.,

$$(a - b) - c = (a - c) - b,$$

etc. Finally $-(-a) = a$ and $a - a = 0$.

REGARDING THE LAWS OF ASSOCIATIVITY. As we saw in Chapter II, Section 6, it is possible, on the basis of the associative law for multiplication, to define the composite products

$$\prod_1^n a_\nu = a_1 a_2 \ldots a_n$$

and to prove their characteristic property

$$\prod_1^m a_\mu \cdot \prod_{\nu=1}^n a_{m+\nu} = \prod_1^{m+n} a_\nu$$

Similarly, we define the sums

$$\sum_1^n a_\nu = a_1 + a_2 \ldots + a_n$$

and prove their characteristic property

$$\sum_1^m a_\mu + \sum_{\nu=1}^n a_{m+\nu} = \sum_1^{m+n} a_\nu$$

By virtue of Ib, the terms of a sum may be interchanged at will, and in commutative rings the same may be done for products.

---

[2] This group is known as the *additive group* of the ring.

REGARDING THE LAWS OF DISTRIBUTIVITY. Whenever the commutative law of multiplication holds, IIIb is, of course, a consequence of IIIa.

By complete induction on n it follows at once from IIIa that

$$a(b_1+b_2+\cdots+b_n)=ab_1+ab_2+\cdots+ab_n,$$

and from IIIb that

$$(a_1+a_2+\cdots+a_n)b=a_1b+a_2b+\cdots+a_nb.$$

Combining these two relations, we obtain the well-known rule for the multiplication of sums:

$$\begin{aligned}(a_1+\cdots+a_n)(b_1+\cdots+b_m)&=a_1b_1+\cdots+a_1b_m\\&\quad+\cdots\cdots\cdots\\&\quad+a_nb_1+\cdots+a_nb_m\\&=\sum_{i=1}^{n}\sum_{k=1}^{m}a_ib_k.\end{aligned}$$

The distributive laws hold for subtraction as well, e.g.,

$$a(b-c)=ab-ac,$$

as can be seen from

$$a(b-c)+ac=a(b-c+c)=ab.$$

In particular

$$a\cdot 0=a(a-a)=a\cdot a-a\cdot a=0,$$

or in words: A *product is zero whenever one of its factors is zero.*

The converse of this theorem is not necessarily true, as will be seen from examples shown below. It may happen that

$$a\cdot b=0,\quad \text{where}\quad a\neq 0,\quad b\neq 0.$$

In such a case a and b are called *zero divisors* or *divisors of zero*, a being a left and b a right zero divisor. (In commutative rings these two definitions coincide.) For the sake of expediency the zero itself is considered a zero divisor. Thus a is called a left zero divisor, if there exists a $b\neq 0$ such that $ab=0$.[3]

If there are no zero divisors in a ring except zero itself, i.e., whenever $ab=0$ implies $a=0$ or $b=0$, the ring is called a *ring without divisors of zero.* If, moreover, the ring is commutative, it is also known as a *domain of integrity.*

EXAMPLES: All examples mentioned in the beginning (ring of integers, ring of rational numbers, etc.) are rings without zero divisors. The ring of the continuous functions in the interval $(-1, +1)$ does have zero divisors; for putting

$$f=f(x)=\max(0, x),$$
$$g=g(x)=\max(0, -x),$$

we have $f\neq 0$[4], $g\neq 0$, $fg=0$.

[3] Assuming that there is at least one element $\neq 0$ in the ring.

[4] $f\neq 0$ means: f is a function distinct from zero, but it does not mean that f does not vanish anywhere.

EXERCISES. 1. The number pairs (a_1, a_2) (a_1, a_2 may be rational numbers) with

$$(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 + b_2),$$
$$(a_1, a_2) \cdot (b_1, b_2) = (a_1 b_1, a_2 b_2)$$

form a ring with zero divisors.

2. It is permissible to cancel a in an equation $ax = ay$, provided a is not a left zero divisor. (In particular, it is possible to cancel any $a \neq 0$ in an integral domain.)

3. Taking as an additive group any arbitrary Abelian group, construct a ring in which the product of any two elements is equal to zero.

THE IDENTITY. If a ring has a left indentity e,

$$ex = x \text{ for every } x,$$

and *simultaneously*, a right identity e',

$$xe' = x \text{ for every } x,$$

then the two identities must be equal, since

$$e = ee' = e'.$$

Then every right identity is equal to e, and so is every left identity. Then e is simply called the identity, and a ring containing such an element is called a *ring with identity*. Frequently, the identity is denoted by 1, although it has to be distinguished from the number 1.

The integers form a ring C with identity; the even numbers, a ring without identity. There are also rings with one or more right identities but without a left identity, or vice versa.

THE INVERSE ELEMENT. If a is an arbitrary element of a ring with identity e, then, by a *left inverse* of a we shall understand an element $a_{(l)}^{-1}$ such that

$$a_{(l)}^{-1} a = e,$$

and by a *right inverse* an element $a_{(r)}^{-1}$ such that

$$a a_{(r)}^{-1} = e.$$

If an element a has both a left and a right inverse, then both are equal, since

$$a_{(l)}^{-1} = a_{(l)}^{-1} (a a_{(r)}^{-1}) = (a_{(l)}^{-1} a) a_{(r)}^{-1} = a_{(r)}^{-1}$$

and hence each right inverse as well as each left inverse of a is equal to this one. In this case we say: *a possesses an inverse element*, and we denote the inverse by a^{-1}.

POWERS AND MULTIPLES. We saw already in Chapter II that, by virtue of the associative law, the powers a^n (n is a positive integer) may be defined for every ring element a, and that the well-known rules hold:

$$\text{(1)} \qquad \begin{cases} a^n \cdot a^m = a^{n+m}, \\ (a^n)^m = a^{nm}, \\ (ab)^n = a^n b^n, \end{cases}$$

the last equality holding for commutative rings.

If the ring has the identity, and if a has an inverse, we may introduce the 0-th and negative powers (Section 6); the rules (1) remain valid.

Since any ring is an additive group, the multiples

$$n \cdot a \qquad (= a + a + \cdots + a, \text{ with } n \text{ terms})$$

may be defined, and we have:

$$(2) \qquad \begin{cases} na + ma = (n+m)a, \\ n \cdot ma = nm \cdot a, \\ n(a+b) = na + nb, \\ n \cdot ab = na \cdot b = a \cdot nb. \end{cases}$$

If we define

$$(-n) \cdot a = -na,$$

just as for powers, the rules (2) will hold for all integral n's and m's (positive, negative, or zero).

The expression $n \cdot a$ should not be regarded as a real product of two ring elements; for, in general, n is not a ring element, but something introduced from the outside: an integer. However, if the ring has the identity e, na may be written as a real product, viz.

$$na = n \cdot ea = ne \cdot a.$$

EXERCISES. 4. A left zero divisor has no left inverse; a right zero divisor has no right inverse. In particular, the zero element has neither a right nor a left inverse. A trivial exception is the ring consisting of but one element 0 which, simultaneously, is the identity and its own inverse.

5. Prove for arbitrary commutative rings the *binomial theorem:*

$$(a+b)^n = a^n + \binom{n}{1} a^{n-1}b + \binom{n}{2} a^{n-2}b^2 + \cdots + b^n,$$

by complete induction on n. Here $\binom{n}{k}$ expresses the integer

$$\frac{n(n-1)\cdots(n-k+1)}{1 \cdot 2 \cdots k} = \frac{n!}{(n-k)!\,k!}$$

6. In a ring with exactly n elements we have for every a:

$$n \cdot a = 0.$$

[Cf. Section 8, where $a^n = e$ was proved.]

7. If a commutes with b, i.e. if $ab = ba$, then a also commutes with $-b$, with nb, and with b^{-1}. If a commutes with b and c, then a also commutes with $b+c$ and with bc.

FIELDS. A ring is called a *skew field*,[5] if

a) it contains at least one element distinct from zero,

b) there is always a solution for the equations

$$(3) \qquad \begin{cases} ax = b, \\ ya = b \end{cases}$$

for $a \neq 0$.

[5] Some authors extend the term "field" to all skew fields and make a distinction between commutative and non-commutative fields.

If, in addition, this ring is commutative, it is simply called a *field* or a *domain of rationality*, sometimes a *commutative field.*

From a) and b) we prove, just as for groups (Chapter II):

c) the existence of a left identity e; for solve the equation $xa=a$ for any $a \neq 0$, and call the solution e. For an arbitrary b, solve $ax=b$; it follows that

$$eb=eax=ax=b.$$

Similarly, the existence of a right identity may be proved, and thus we have proved the *existence of an identity.*

Moreover, from b) follows at once:

d) the existence of a left inverse a^{-1} for every $a \neq 0$. As in the case of groups, it can be shown that the left inverse a^{-1} is at the same time a right inverse.

As in the case of groups, we may show that, *conversely*, b) *follows from* c) *and* d).

EXERCISE. 8. Carry out the proof of the above.

A skew field has no zero divisors; for by multiplying $ab=0, a \neq 0$ by a^{-1}, it follows at once that $b=0$.

Equations (3) have unique solutions; for from the existence of two solutions x, x', say, of the first equation it would follow that

$$ax=ax',$$

and upon multiplying by a^{-1} on the left:

$$x=x'.$$

The solutions of (3), of course, are

$$x=a^{-1}b,$$
$$y=ba^{-1}.$$

In the commutative case we have $a^{-1}b=ba^{-1}$; for this we may write $\frac{b}{a}$.

In a skew field the non-zero elements form a group under multiplication, viz. the multiplicative group of the skew field.

Thus, a skew field unites in itself two groups; the multiplicative and the additive groups. They are connected by the laws of distributivity.

EXAMPLES. 1. The rational numbers, the real numbers, the complex numbers form commutative fields.

2. A field containing only two elements 0 and 1 may be constructed as follows: Multiply the elements like the numbers 0 and 1. For addition, let 0 be the zero element:

$$0+0=0,\ 0+1=1+0=1;$$

furthermore, let $1+1=0$. The rule of addition is the same as the rule of combination for a cyclic group of two elements (Section 7); thus the laws of addition hold. The laws of multiplication hold, since they are valid for the ordinary numbers

0 and 1. The first distributive law is proved by enumerating all possibilities. As soon as a zero occurs in it, it becomes trivial; therefore, we need verify only

$$1 \cdot (1+1) = 1 \cdot 1 + 1 \cdot 1,$$

which yields $0 = 0$. Finally, the equation $1 \cdot x = a$ has a solution for every a; the solution is: $x = a$.

EXERCISES. 9. Construct a field of three elements. [First discuss what structures the additive and the multiplicative groups may have.]

10. An integral domain with a finite number of elements is a field. (Cf. the corresponding theorem for groups in Chapter II, Section 6.)

12. HOMOMORPHISM AND ISOMORPHISM

Let $\mathfrak{S}$, $\bar{\mathfrak{S}}$ be systems of double composition, and let $\mathfrak{S}$ be mapped upon $\bar{\mathfrak{S}}$ so that with every a of $\mathfrak{S}$ there is associated a $\bar{a}$ of $\bar{\mathfrak{S}}$, and so that, conversely, every $\bar{a}$ of $\bar{\mathfrak{S}}$ is associated with at least one a of $\mathfrak{S}$. The mapping is called a *homomorphism* (or *many-one isomorphism*), if $a \to \bar{a}$ and $b \to \bar{b}$ imply

$$a + b \to \bar{a} + \bar{b}$$

and

$$a \cdot b \to \bar{a} \cdot \bar{b}.$$

$\bar{\mathfrak{S}}$ is then called a *homomorphic image* of $\mathfrak{S}$; in symbols $\mathfrak{S} \sim \bar{\mathfrak{S}}$.

If, moreover, the mapping constitutes a one-to-one correspondence, i.e. if every $\bar{a}$ belongs to exactly one a, then the mapping is called an *isomorphism* (or: 1-*isomorphism* or *one-to-one isomorphism*), and it is symbolized by $\mathfrak{S} \cong \bar{\mathfrak{S}}$. The systems $\mathfrak{S}$, $\bar{\mathfrak{S}}$ are then called isomorphic. The relation $\mathfrak{S} \cong \bar{\mathfrak{S}}$ is reflexive, transitive, and, since the inverse mapping of an isomorphism is again an isomorphism, also symmetric.

The homomorphic image of a ring is itself a ring.

PROOF: Let $\mathfrak{R}$ be a ring, $\bar{\mathfrak{R}}$ a system of double composition, and $\mathfrak{R} \sim \bar{\mathfrak{R}}$. We have to show that $\bar{\mathfrak{R}}$ is itself a ring. The proof is the same as that for groups (Section 10):

If $\bar{a}$, $\bar{b}$, $\bar{c}$ are any three elements of $\bar{\mathfrak{R}}$, and if any rule of operation is to be proved, say $\bar{a}(\bar{b}+\bar{c}) = \bar{a}\bar{b} + \bar{a}\bar{c}$, then three inverse images (pre-images) a, b, c can be found for $\bar{a}$, $\bar{b}$, $\bar{c}$. Since $\mathfrak{R}$ is a ring, $a(b+c) = ab + ac$, whence it follows that $\bar{a}(\bar{b}+\bar{c}) = \bar{a}\bar{b} + \bar{a}\bar{c}$ because of the homomorphism. The same procedure may be applied to all laws of associativity, commutativity, and distributivity. In order to prove the solvability of the equation $\bar{a} + \bar{x} = \bar{b}$, we again take inverse images a, b, solve $a + x = b$, and then, because of the homomorphism, we have $\bar{a} + \bar{x} = \bar{b}$.

Under a homomorphism, the zero element and the inverse in $\bar{\mathfrak{R}}$ correspond, respectively, to the zero element 0 of $\mathfrak{R}$ and the inverse $-a$ of any element a. If $\mathfrak{R}$ has an identity e, then the identity in $\bar{\mathfrak{R}}$ corresponds to it.

PROOF, as in the case of groups:

$$\begin{aligned} a+0 &= a \text{ implies } \bar{a}+\bar{0} &&= \bar{a};\\ -a+a &= 0 \text{ implies } -\bar{a}+\bar{a} &&= \bar{0};\\ ae &= a \text{ implies } \bar{a}\bar{e} &&= \bar{a}. \end{aligned}$$

Evidently, if $\mathfrak{R}$ is commutative, so is $\overline{\mathfrak{R}}$.

If $\mathfrak{R}$ is an integral domain, $\overline{\mathfrak{R}}$ need not be one, as we shall see afterwards: similarly $\overline{\mathfrak{R}}$ may be an integral domain without $\mathfrak{R}$ being one. However, if we are dealing with an isomorphic mapping, then, of course, all algebraic properties are carried over from $\mathfrak{R}$ into $\overline{\mathfrak{R}}$, whence it follows that *the isomorphic image of an integral domain or of a field is, respectively, an integral domain or a field.*

A theorem which apparently is trivial in this connection, but which will prove very useful in the sequel, is the following:

Let $\mathfrak{R}$ and $\mathfrak{S}'$ be two mutually exclusive rings. Let $\mathfrak{S}'$ contain a subring $\mathfrak{R}'$ 1-isomorphic with $\mathfrak{R}$. Then there exists a ring $\mathfrak{S} \cong \mathfrak{S}'$ which includes $\mathfrak{R}$ itself.

PROOF: From $\mathfrak{S}'$ we eliminate the elements of $\mathfrak{R}'$ and replace them by the elements of $\mathfrak{R}$ to which they correspond under the isomorphism. Now, we define the sums and products for the unreplaced and replaced elements so that they correspond exactly to the sums and products in $\mathfrak{S}'$. (If, e.g., before the replacement $a'b'=c'$, and if a' is replaced by a, while b' and c' remain unaltered, we define: $ab'=c'$.) In this fashion we obtain from $\mathfrak{S}'$ a ring $\mathfrak{S} \cong \mathfrak{S}'$ which in fact includes $\mathfrak{R}$.

13. THE CONCEPT OF A FIELD OF QUOTIENTS

If a commutative ring $\mathfrak{R}$ is embedded in a skew field Ω we may form quotients in Ω from the elements of $\mathfrak{R}$ thus[6]:

$$\frac{a}{b} = a\,b^{-1} = b^{-1}\,a \quad (b \neq 0)$$

For them the following rules of operation hold:

$$\text{(1)} \qquad \begin{cases} \dfrac{a}{b} = \dfrac{c}{d} \text{ when, and only when, } ad = bc;\\[2ex] \dfrac{a}{b} + \dfrac{c}{d} = \dfrac{ad+bc}{bd};\\[2ex] \dfrac{a}{b} \cdot \dfrac{c}{d} = \dfrac{ac}{bd}. \end{cases}$$

For if we multiply both sides of each equation by bd, we get the same result in each case; and $bdx = bdy$ implies $x = y$.

Thus we can see that the quotients $\frac{a}{b}$ form a commutative field P, called the *field of quotients* of the commutative ring $\mathfrak{R}$. Furthermore, it can be seen from (1) that the manner in which fractions are compared, added, and multiplied, will be known, whenever these operations can be performed on their numerators and denominators, i.e., on the elements of $\mathfrak{R}$; in other words, the structure of the field of

[6] From $ab = ba$ follows $ab^{-1} = b^{-1}a$ on multiplying by b^{-1} on the left and on the right.

quotients P is completely determined by that of $\mathfrak{R}$, or: *Fields of quotients of isomorphic rings are isomorphic.* In particular, any two fields of quotients of a single ring are always isomorphic, or: *The field of quotients* P *is, except for isomorphism, uniquely determined by the ring* $\mathfrak{R}$, *provided there exists at all a field of quotients for the ring* $\mathfrak{R}$.

Now the following question arises: What commutative rings possess a field of quotients, or, which is the same thing, what commutative rings can be embedded in a field?

In order that a ring $\mathfrak{R}$ may be embedded in a field, it is first necessary that there be no zero divisors in $\mathfrak{R}$; for a field has no zero divisors. In the commutative case this condition is also sufficient: *Any integral domain* $\mathfrak{R}$ *can be embedded in a field.*[7]

PROOF: We may disregard the trivial case where $\mathfrak{R}$ consists only of a zero element. Let us consider the set of all pairs of elements (a, b), where $b \neq 0$. With these pairs we shall later associate fractions $\frac{a}{b}$.

Let $(a, b) \sim (c, d)$ if $ad = bc$. [Cf. the above formulae (1).] The relation $\sim$ thus defined is obviously reflexive and symmetric; it is likewise transitive, since from

$$(a, b) \sim (c, d), \qquad (c, d) \sim (e, f)$$

it follows that

$$ad = bc, \qquad cf = de,$$

hence

$$adf = bcf = bde,$$

This implies, since $d \neq 0$ and $\mathfrak{R}$ is commutative,

$$af = be,$$

$$(a, b) \sim (e, f).$$

Consequently, the relation $\sim$ has all the properties of an equivalence relation and therefore defines (according to Chapter I, Section 5) a partition for the pairs (a, b), equivalent pairs being classified in the equivalence class. Let the equivalence class containing (a, b) be symbolized by $\frac{a}{b}$. According to this definition $\frac{a}{b} = \frac{c}{d}$ when, and only when, $(a, b) \sim (c, d)$, i.e., when $ad = bc$.

In accordance with the earlier formula (1) we now define the sum and the product of the new symbols $\frac{a}{b}$ by:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}, \tag{2}$$

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}. \tag{3}$$

The definitions are admissible for the following reasons: First, $bd \neq 0$ if $b \neq 0$ and $d \neq 0$; $\frac{ad + bc}{bd}$ and $\frac{ac}{bd}$ are therefore permissible symbols: secondly, the right-hand sides are independent of the choice of the representatives (a, b) and

[7] For non-commutative rings without zero divisors this theorem does not hold any longer; cf. A. Malcev in *Math. Ann.* 113 (1936).

(c, d) of the classes $\frac{a}{b}$ and $\frac{c}{d}$; for replacing a and b in (2) by a' and b', where

$$ab' = ba',$$

we find

$$adb' = a'db,$$
$$adb' + bcb' = a'db + b'cb,$$
$$(ad + bc)b'd = (a'd + b'c)bd,$$

hence

$$\frac{ad + bc}{bd} = \frac{a'd + b'c}{b'd}$$

Similarly:

$$ab' = ba',$$
$$acb'd = a'cbd,$$
$$\frac{ac}{bd} = \frac{a'c}{b'd}.$$

Similar results will be obtained when (c, d) is replaced by (c', d'), where $cd' = dc'$.

We can show without difficulty that all field properties are fulfilled. The associative law of addition, for example, is obtained thus:

$$\frac{a}{b} + \left(\frac{c}{d} + \frac{e}{f}\right) = \frac{a}{b} + \frac{cf + de}{df} = \frac{adf + bcf + bde}{bdf}$$
$$\left(\frac{a}{b} + \frac{c}{d}\right) + \frac{e}{f} = \frac{ad + bc}{bd} + \frac{e}{f} = \frac{adf + bcf + bde}{bdf},$$

and the other laws are obtained in the same way.

Evidently, the field constructed is commutative. If it is to include the ring $\mathfrak{R}$, certain fractions have to be identified with elements of $\mathfrak{R}$. This is done as follows:

We associate all fractions $\frac{cb}{b}$ with the element c, where $b \neq 0$. These fractions are all equal:

$$\frac{cb}{b} = \frac{cb'}{b'}, \text{ since } (cb)b' = b(cb').$$

Thus, with every element c there is associated only *one* fraction, and distinct fractions are associated with distin t elements c, c'; for it follows from

$$\frac{cb}{b} = \frac{c'b'}{b'}$$

that

$$cbb' = bc'b'$$

or, since $b \neq 0$, $b' \neq 0$, we may divide each member by bb' and have:

$$c = c'.$$

Therefore, these fractions are in a one-to-one correspondence with the elements of $\mathfrak{R}$.

If $c_1 + c_2 = c_3$ or $c_1 c_2 = c_3$ are in $\mathfrak{R}$, it follows for arbitrary $b_1 \neq 0$, $b_2 \neq 0$, and $b_3 = b_1 b_2$ that, respectively

$$\frac{c_1 b_1}{b_1} + \frac{c_2 b_2}{b_2} = \frac{c_1 b_1 b_2 + c_2 b_1 b_2}{b_1 b_2} = \frac{c_3 b_3}{b_3}$$

and

$$\frac{c_1 b_1}{b_1} \cdot \frac{c_2 b_2}{b_2} = \frac{c_1 c_2 b_1 b_2}{b_1 b_2} = \frac{c_2 b_2}{b_2}.$$

The associated fractions $\frac{c_i b_i}{b_i}$ thus add and multiply exactly as the ring elements c_i; they form a domain isomorphic with $\mathfrak{R}$. Consequently, we may replace the fractions $\frac{cb}{b}$ by the corresponding elements c (end of Section 12). We thus find a field including the ring $\mathfrak{R}$.

Thus, the existence of a comprehending field for every integral domain $\mathfrak{R}$ has been proved.

The formation of quotients is the first tool we employ for obtaining rings or fields from other rings. Thus, for example, from the ring C of the ordinary integers the field Γ of rational numbers is derived by this method.

EXERCISE. Show that any commutative ring (with or without a zero divisor) can be embedded in a "quotient ring" consisting of all quotients $\frac{a}{b}$, where b takes the values of all non-zero divisors. Generally speaking, we can let b take the values of any set $\mathfrak{M}$ of non-zero divisors which contains for any two elements b_1, b_2 their product $b_1 b_2$, and in this way a quotient ring $\mathfrak{R}_{\mathfrak{M}}$ is obtained.

14. VECTOR SPACES AND HYPERCOMPLEX SYSTEMS (LINEAR ASSOCIATIVE ALGEBRAS)

Let $\mathfrak{R}$ be a ring with an identity, and let the elements of the ring be denoted by Greek letters $\alpha, \beta, \ldots\ldots$ Let $\mathfrak{G}$ be an additive Abelian group. Its elements will be designated by the italics $u, v, \ldots$

The module $\mathfrak{G}$ is called an *n-termed module of linear forms* or *an n-dimensional vector space* if, besides addition in $\mathfrak{G}$, there is defined a multiplication of the elements of $\mathfrak{R}$ by the elements of $\mathfrak{G}$ such that

1. the product αu of an element α of $\mathfrak{R}$ and an element u of $\mathfrak{G}$ always belongs to $\mathfrak{G}$.

2. $\alpha(u + v) = \alpha u + \alpha v$.

3. $(\alpha + \beta) u = \alpha u + \beta u$.

4. $(\alpha\beta) u = \alpha(\beta u)$.

5. All elements of $\mathfrak{G}$ are uniquely expressible as linear forms $\alpha_1 u_1 + \cdots + \alpha_n u_n$ by means of n fixed "basis elements" $u_1, \ldots, u_n$.

It follows from 2. and 4. that

$$\beta(\alpha_1 u_1 + \cdots + \alpha_n u_n) = (\beta\alpha_1) u_1 + \cdots + (\beta\alpha_n) u_n. \tag{1}$$

If, in particular, we set $\beta = 1$ and $\alpha_1 u_1 + \cdots + \alpha_n u_n = u$, it follows that

$$1 \cdot u = u. \tag{2}$$

Furthermore, it follows from 3. that

$$\begin{aligned}(\alpha_1 u_1 + \cdots + \alpha_n u_n) + (\beta_1 u_1 + \cdots + \beta_n u_n) \\ = (\alpha_1 + \beta_1) u + \cdots + (\alpha_n + \beta_n) u_n.\end{aligned}$$

Every element $u = \alpha_1 u_1 + \ldots + \alpha_n u_n$ of the vector space is uniquely represented by an ordered set of n elements $(\alpha_1, \ldots, \alpha_n)$, called the *components* of u (with respect to the basis $u_1, \ldots, u_n$). The number n of the basis elements or of the components is called the dimension of the vector space. The addition of two elements of $\mathfrak{G}$ is performed by adding their components according to (3); the multiplication by β is performed by multiplying all the components by β according to (1).

Thus, except for isomorphism, the vector space is uniquely determined by the ring $\mathfrak{R}$ and the dimension n.

On the basis of this theorem, one can, for a given $\mathfrak{R}$, take any vector space of a given dimension as a model for all. This construction is carried out most conveniently by regarding a vector as an ordered set of n elements $\alpha_1, \ldots, \alpha_n$ of $\mathfrak{R}$. The sum of two vectors $(\alpha_1, \ldots, \alpha_n)$ and $(\beta_1, \ldots, \beta_n)$ is defined by $(\alpha_1 + \beta_1, \ldots, \alpha_n + \beta_n)$, and the product $\beta(\alpha_1, \ldots, \alpha_n)$ by $(\beta\alpha_1, \ldots, \beta\alpha_n)$. Then the rules of operation 1. to 4. are fulfilled automatically. If we now put

$$(1, 0, \ldots, 0) = u_1$$
$$(0, 1, \ldots, 0) = u_2$$
$$\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots$$
$$(0, 0, \ldots, 1) = u_n,$$

we have

$$(\alpha_1, \ldots, \alpha_n) = (\alpha_1, 0, \ldots, 0) + (0, \alpha_2, \ldots, 0) + \cdots + (0, 0, \ldots, \alpha_n)$$
$$= \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n,$$

hence 5. is satisfied as well. Thus it is seen that the vectors $(\alpha_1, \ldots, \alpha_n)$ actually form an n-dimensional vector space in the sense of our definition.

A vector space $\mathfrak{G}$ becomes a ring if, in addition, we define multiplication for the elements $u, v, \ldots$. Besides the associative law and the two distributive laws the following property is postulated for this multiplication:

(4) $$(\alpha u)\, v = u\,(\alpha v) = \alpha\,(uv) \quad \text{for} \quad \alpha \in \mathfrak{R}.$$

$$(\alpha u)\,(\beta v) = (\alpha\beta)\,(uv)$$

and, furthermore,

(5) $$\Big(\sum_j \alpha_j u_j\Big)\Big(\sum_k \beta_k u_k\Big) = \sum_j \sum_k \alpha_j \beta_k\,(u_j u_k).$$

Hence the products uv are all computable, as soon as the products $u_j u_k$ are known. Of course, these products will be linear combinations of $u_1, \ldots, u_n$:

(6) $$u_j u_k = \sum_l \gamma_{jk}^l u_l.$$

According to (5) and (6), multiplication in a hypercomplex system is completely determined by n^3 constants γ_{jk}^l. They are called the structure constants of the system.

Let a ring $\mathfrak{R}$, a vector space $\mathfrak{G}$ over $\mathfrak{R}$, and a system of structure constants γ_{jk}^l be given. Now, if multiplication is defined by (5) and (6), the distributive

laws will always hold. If $\mathfrak{R}$ is commutative, (4) is also satisfied. However, the associative law of multiplication is not fulfilled automatically. For arbitrary sums

$$u = \sum_j \alpha_j u_j, \qquad v = \sum_k \beta_k u_k, \qquad w = \sum_l \gamma_l u_l,$$

the associative law holds as soon as it holds for the products $u_j u_k u_l$ of the basis elements, i.e., if

$$u_j (u_k u_l) = (u_j u_k) u_l . \tag{7}$$

(7) is an additional requirement for the structure constants γ_{jk}^l. If this condition is satisfied, the formulae (1), (3), (5), (6) define the operations of a hypercomplex system over the ring $\mathfrak{R}$.

If $\mathfrak{R}$ is commutative, and if the multiplication of the u_j is likewise commutative, i.e. if $u_j u_k = u_k u_j$, then $\mathfrak{G}$ is a commutative ring. If $\mathfrak{G}$ contains an identity e, the multiples αe in $\mathfrak{G}$ form a ring 1-isomorphic with $\mathfrak{R}$; hence they may be identified with the elements α of $\mathfrak{R}$.

If $\mathfrak{G}$ is a skew field, $\mathfrak{G}$ is also called a *division algebra.*

First example: *Definition of the ordinary complex numbers.*

Choose as $\mathfrak{R}$ the field of real numbers, and let the multiplication of the basis elements e and i of a two-dimensional vector space $\mathfrak{G}$ be defined by

$$e \cdot e = e \qquad e \cdot i = i$$
$$i \cdot e = i \qquad i \cdot i = -e .$$

The multiplication of e and i is associative and commutative; hence $\mathfrak{G}$ becomes a commutative ring with the identity e. The multiples αe may be identified with the real numbers α. Thus we write $a + bi$ instead of $ae + bi$. Since

$$(a - bi)(a + bi) = a^2 + b^2 > 0 \text{ (except for } a = b = 0),$$

every element $a + bi$ distinct from zero has an inverse, viz. $(a^2 + b^2)^{-1}(a - bi)$. Therefore, the ring $\mathfrak{G}$ is a field, the field of complex numbers.

The same is true if we choose as $\mathfrak{R}$ the field of rational numbers. The system $\mathfrak{G}$ thus formed is called *Gaussian number field.* If we take for $\mathfrak{R}$ the ring of integers, then $\mathfrak{G}$ becomes the *ring of the Gaussian integers* $a + bi$.

Second example. *The algebra of quaternions.* Let $\mathfrak{R}$ again be the field of real or rational numbers, and $\mathfrak{G}$ a four-dimensional vector space with basis elements e, j, k, l. The rules of multiplication

$$ee = e; \qquad jj = kk = ll = -e;$$
$$ej = je = j; \qquad ek = ke = k; \qquad el = le = l;$$
$$jk = l; \qquad kj = -l;$$
$$kl = j; \qquad lk = -j$$
$$lj = k; \qquad jl = -k$$

are associative, but not commutative. Hence we obtain a non-commutative ring with the identity e. The multiples ae may be identified with the numbers a. The elements $a + bj + ck + dl$ are called quaternions. Since

$$(a - bj - ck - dl)(a + bj + ck + dl) = a^2 + b^2 + c^2 + d^2$$

every quaternion $a + bj + ck + dl$ distinct from zero has an inverse

$$(a^2 + b^2 + c^2 + d^2)^{-1}(a - bj - ck - dl).$$

Thus, the system of quaternions is a division algebra, called the *algebra of quaternions.*

Third example: *The group ring of a finite group.* If we choose elements of a finite group $\mathfrak{g}$ as the basis elements of a vector space $\mathfrak{R}_\mathfrak{g}$, the associative law (7) is fulfilled automatically. The hypercomplex system $\mathfrak{R}_\mathfrak{g}$ thus formed is called the *group ring* of $\mathfrak{g}$ over $\mathfrak{R}$.

EXERCISES. 1. If we define a matrix of the n-th degree as a system of n^2 elements $\alpha_{jk}(j = 1, \ldots, n, k = 1, \ldots, n)$ of the ring $\mathfrak{R}$, and if we define the sum (σ_{jk}) and the product (π_{jk}) of two matrices (α_{jk}), (β_{jk}) in the customary manner by

$$\sigma_{jk} = \alpha_{jk} + \beta_{jk}$$

$$\pi_{jl} = \sum_{k=1}^{n} \alpha_{jk}\beta_{kl},$$

then the matrices of the n-th degree form a hypercomplex system of rank n^2 over $\mathfrak{R}$.

2. The two-rowed complex matrices

$$\begin{pmatrix} a + ib & c + id \\ -c + id & a - ib \end{pmatrix}$$

form a linear associative algebra isomorphic to the quaternion field. (The proof of this fact is at the same time a simple proof of the law of associativity for the multiplication of quaternions.)

The generalization of the hypercomplex systems to systems of infinite rank is immediate. Consider an infinite number of basis elements $u_1, u_2, \ldots$, for which multiplication according to (6) and (7) is defined. We consider only *finite* sums $\sum \alpha_j u_j$ as elements af the generalized hypercomplex system $\mathfrak{S}$. All considerations in this paragraph hold for such "infinite linear algebras" as well.

15. POLYNOMIAL RINGS

Let $\mathfrak{R}$ be a ring. With a new symbol x not belonging to $\mathfrak{R}$ we form the expressions

$$f(x) = \sum a_\nu x^\nu$$

in which the sum is taken over a finite number of different integers $\nu \geqq 0$, and where the "*coefficients*" a_ν belong to the ring $\mathfrak{R}$; e.g.:

$$f(x) = a_0x^0 + a_3x^3 + a_5x^5$$

These expressions are called *polynomials;* the symbol x is called an *indeterminate.* Thus, an indeterminate is only a symbol, a letter, nothing else. Two polynomials

are called equal, only when they contain exactly the same terms, aside from terms with zero coefficients, which may be omitted or included at will.

If we add or multiply polynomials $f(x)$ and $g(x)$ according to the rules of high school algebra, assuming that all powers of x commute with the ring elements ($ax^\nu = x^\nu a$) and collect terms with the same power of x, we obtain a polynomial $\sum c_\nu x^\nu$. In case of addition we have

$$c_\nu = a_\nu + b_\nu \tag{1}$$

(taking $a_\nu = 0$ or $b_\nu = 0$ if a_ν or b_ν is missing), and in case of multiplication we have

$$c_\nu = \sum_{\sigma+\tau=\nu} a_\sigma b_\tau . \tag{2}$$

We now *define* sum and product of two polynomials by formulae (1) and (2) and assert:

The polynomials form a ring.

The properties of addition are obvious, since addition of polynomials is reduced to the addition of their coefficients. The first distributive law follows from

$$\sum_{\sigma+\tau=\nu} a_\sigma (b_\tau + c_\tau) = \sum_{\sigma+\tau=\nu} a_\sigma b_\tau + \sum_{\sigma+\tau=\nu} a_\sigma c_\tau$$

and the second is obtained in a similar way. Finally, the associative law of multiplication is obtained from

$$\sum_{\alpha+\tau=\nu} a_\alpha \Big(\sum_{\beta+\gamma=\tau} b_\beta c_\gamma\Big) = \sum_{\alpha+\beta+\gamma=\nu} a_\alpha b_\beta c_\gamma ,$$

$$\sum_{\varrho+\gamma=\nu} \Big(\sum_{\alpha+\beta=\varrho} a_\alpha b_\beta\Big) c_\gamma = \sum_{\alpha+\beta+\gamma=\nu} a_\alpha b_\beta c_\gamma .$$

The polynomial ring derived from $\mathfrak{R}$ is denoted by $\mathfrak{R}[x]$.

An objection against the here given definition of polynomials might be that the summation symbol is used in

$$\sum_\nu a_\nu x^\nu ,$$

even before addition is defined. In order to meet this objection, we might first define a polynomial as just a finite set of coefficients with indices, such as $(a_0, \ldots, a_\lambda, \ldots, a_n)$, and then define sum and product by (1) and (2). After that we may, on the basis of the sum definition, represent any polynomial as the sum of monomials (a_ν), which we may also call $a_\nu x^\nu$, thus giving the symbol $\sum$ in $\sum a_\nu x^\nu$ the meaning of actual summation.

If $\mathfrak{R}$ possesses an identity, the polynomial ring $\mathfrak{R}[x]$ may also be defined as an infinite hypercomplex system with basis elements $x^0, x^1, x^2, \ldots$. From a didactical viewpoint, however, the introduction of the polynomial concept given at the outset is preferable.

The *degree* of a polynomial distinct from zero is the greatest number ν for which $a_\nu \neq 0$. This a_ν is known as the *leading* or the *highest coefficient.*

Polynomials of degree zero are of the form $a_0 x^0$. These polynomials may be identified with the elements a_0 of the original ring $\mathfrak{R}$. This identification is

permissible, since these special polynomials form a system 1-isomorphic with the original ring $\mathfrak{R}$ (cf. end of Section 12). Thus the polynomial ring $\mathfrak{R}[x]$ includes $\mathfrak{R}$.

The transition from $\mathfrak{R}$ to $\mathfrak{R}[x]$ is also known as the *adjunction* (in this case ring adjunction) *of an indeterminate* x. If $\mathfrak{R}$ is commutative, so is $\mathfrak{R}[x]$.

If we adjoin to the ring $\mathfrak{R}$ the indeterminates $x_1, \ldots, x_n$, successively, i.e. if we form $\mathfrak{R}[x_1][x_2]\ldots[x_n]$, we obtain a polynomial ring $\mathfrak{R}[x_1, \ldots, x_n]$, consisting of all the sums

$$\sum a_{\alpha_1 \ldots \alpha_n} x_1^{\alpha_1} \ldots x_n^{\alpha_n}.$$

By changing everywhere in such a polynomial the order of the factors $x_1^{\alpha}, \ldots, x_n^{\alpha_n}$, the polynomial ring $\mathfrak{R}[x_1][x_2]\ldots[x_n]$ may be identified with the polynomial ring of the interchanged indeterminates, say with $\mathfrak{R}[x_2][x_n]\ldots[x_1]$. This identification is permitted, since the interchanging of the x_i has no effect on the definition of sum and product. $\mathfrak{R}[x_1, \ldots, x_n]$ is called the *polynomial ring in the n indeterminates* $x_1, \ldots, x_n$.

The reason why polynomials are so extremely useful is that we may substitute for the indeterminate x any element of $\mathfrak{R}$ or even of a larger ring, without destroying the validity of the relations $f+g=h$ and $f \cdot g = k$, provided that the element α which is substituted for x commutes with all elements of $\mathfrak{R}$. This results from the fact that we defined the sum h and the product k by the ordinary laws of high school algebra; in fact, these laws remain valid when x is replaced by any ring element α; only α must commute with all coefficients a, since we assumed $a x^{\nu} = x^{\nu} a$. The ring element arising from $f(x)$ by substituting $x=\alpha$ is denoted by $f(\alpha)$.

For polynomials in several variables the same possibility of substitution exists, provided the elements which are substituted for $x_1, \ldots, x_n$ commute mutually and commute with the elements of $\mathfrak{R}$. In commutative rings we may substitute whatever elements we wish.

Because of this possibility of substitution the polynomials are also called *rational integral functions of the variables* $x_1, \ldots, x_n$.

In polynomials over the ring of integers without a constant term the substitution can be extended even further; we may substitute an element of any given ring for x, whether or not the ring includes the ring of integers.

If $\mathfrak{R}$ *is an integral domain, so is* $\mathfrak{R}[x]$.

PROOF. If $f(x) \neq 0$ and $g(x) \neq 0$, and if a_{α} is the leading coefficient (distinct from zero) in $f(x)$, and, similarly, if b_{β} is the leading coefficient in $g(x)$. then $a_{\alpha} b_{\beta} \neq 0$ is the coefficient of $x^{\alpha+\beta}$ in $f(x) \cdot g(x)$; hence $f(x) \cdot g(x) \neq 0$. Therefore there are no divisors of zero.

From the proof we infer the *corollary*. If $\mathfrak{R}$ is an *integral domain, the degree* $f(x) \cdot g(x)$ *is the sum of the degrees of* $f(x)$ *and* $g(x)$.

For polynomials in n variables we have by complete induction:

If $\mathfrak{R}$ *is an integral domain, so is* $\mathfrak{R}[x_1, \ldots, x_n]$.

By the degree of a term $a_{\alpha_1 \ldots \alpha_r} x_1^{\alpha_1} \ldots x_r^{\alpha_r}$ we mean the sum of the exponents $\sum \alpha_i$. By the degree of a non-vanishing polynomial we mean the highest of the degrees of the terms distinct from zero. A polynomial is said to be *homogeneous* [8] or to be a *form* if all of its terms are of the same degree. Products of homogeneous polynomials are themselves homogeneous, and the degree of the product is equal to the sum of the degrees of the factors if $\mathfrak{R}$ is an integral domain.

Non-homogeneous polynomials may be expressed (uniquely) as sums of homogeneous constitutents of various degrees. If we multiply two such polynomials f, g of degrees m and n respectively, then the product of the homogeneous constituents of highest degree is, in case of an integral domain $\mathfrak{R}$, a non-vanishing form of degree $m+n$. All the other constituents of $f \cdot g$ are of lower degree; hence the degree of $f \cdot g$ is again $m+n$. Hence the above mentioned corollary is also valid for polynomials in any number of indeterminates.

THE DIVISION ALGORITHM. If $\mathfrak{R}$ is a ring with the identity 1, and if, furthermore,

$$g(x) = \sum c_\nu x^\nu$$

is a polynomial with the leading coefficient $c_n = 1$, and if

$$f(x) = \sum a_\nu x^\nu$$

is an arbitrary polynomial of degree $m \geqq n$, we can make the leading coefficient a_m vanish by subtracting from f a multiple of g, namely $a_m x^{m-n} g$. If after this subtraction the degree is still $\geqq n$, we can again remove the leading coefficient by subtracting another multiple of g. Continuing in this way, we eventually lower the degree of the remainder to less than n and we have:

$$f - qg = r, \tag{3}$$

where r is of lower degree than g, or equal to zero. This process is known as the *division algorithm.*

If, in particular, $\mathfrak{R}$ is a field and $g \neq 0$, then the assumption that $c_n = 1$ is superfluous; for in this case, one may, if necessary, multiply g by c_n^{-1} and thus reduce the leading coefficient to 1.

EXERCISE. If $x, y, \ldots$ are an infinite number of symbols, we may consider the totality of all $\mathfrak{R}$-polynomials in these indeterminates. Every polynomial, however, must contain but a finite number of these indeterminates. Prove that the domain thus defined is, respectively, a ring or integral domain whenever $\mathfrak{R}$ is a ring or integral domain.

[8] Translator's Note: British authors frequently employ the word "quantics" for homogeneous polynomials.

16. IDEALS. RESIDUE CLASS RINGS

Let $\mathfrak{o}$ be a ring.

For a subset of $\mathfrak{o}$ to be itself a ring (subring of $\mathfrak{o}$), it is necessary and sufficient

1) that the subset be a subgroup of the additive group; in other words, that it contain with a and b, also $a - b$[9] (*property of the module*),

2) that it contain with a and b, also $a \cdot b$.

Some of the subrings, which we shall call *ideals*, play a special rôle, in analogy to the normal divisors in group theory.

A non-empty subset $\mathfrak{m}$ of $\mathfrak{o}$ is called an *ideal*, or better *a right ideal* if

1. $a \in \mathfrak{m}$ and $b \in \mathfrak{m}$ imply $a - b \in \mathfrak{m}$ (property of the module),

2. $a \in \mathfrak{m}$ implies $ar \in \mathfrak{m}$ for an arbitrary r in $\mathfrak{o}$. In words: the module $\mathfrak{m}$ shall contain all "*right multiples*" $a \cdot r$ for every a.

Similarly, a module is called a *left ideal* if $a \in \mathfrak{m}$ implies $ra \in \mathfrak{m}$ for an arbitrary r in $\mathfrak{o}$.

Finally, $\mathfrak{m}$ is called a *two-sided ideal* if $\mathfrak{m}$ is both a left and a right ideal.

For commutative rings the three concepts coincide, and we speak simply of *ideals. In this section we shall further assume that $\mathfrak{o}$ is a commutative ring.* Ideals will always be denoted by small German letters.

Examples of ideals:

1. *The null ideal*, consisting of the zero element alone.

2. *The unit ideal* $\mathfrak{o}$, including all elements of the ring.

3. *The ideal (a) generated by an element a;* it consists of all expressions of the form

$$ra + na \qquad (r \in \mathfrak{o},\ n \text{ is an integer}).$$

It is easily seen that this set is always an ideal: the difference of two such expressions is obviously of the same form, and an arbitrary multiple is of the form

$$s \cdot (ra + na) = (sr + ns) \cdot a,$$

The ideal (a), evidently, is the smallest ideal containing a: for every such ideal has to contain at least all multiples ra and all sums $\pm \sum a = na$, hence all sums $ra + na$. Therefore, the ideal (a) may also be defined as the intersection of all ideals containing a as an element.

If the ring $\mathfrak{o}$ has the identity e, we may write $ra + nea = (r + ne)a = r'a$ instead of $ra + na$; *in this case (a) thus consists of all ordinary multiples ra.* For example, the ideal (2) in the ring of integers consists of all even integers.

An ideal (a) generated by an element a is called a *principal ideal.* The null ideal (0) is always a principal ideal, and so is the unit ideal $\mathfrak{o}$, provided $\mathfrak{o}$ contains the identity e; then $\mathfrak{o} = (e)$.

[9] From this it follows already that the set contains also the zero and all sums $a + b$; cf. Section 7.

4. Similarly, the ideal generated by several elements $a_1, \ldots, a_n$ may be defined as the totality of all sums of the form

$$\sum r_i a_i + \sum n_j a_j$$

(or, if $\mathfrak{o}$ contains the identity, as $\sum r_i a_i$), or as the intersection of all ideals of $\mathfrak{o}$ containting the elements $a_1, \ldots, a_n$. The ideal is denoted by $(a_1, \ldots, a_n)$, and $a_1, \ldots, a_n$ are said to form an *ideal basis.*

5. In a similar manner we can define the ideal $(\mathfrak{M})$ generated by an infinite set $\mathfrak{M}$; it is the totality of all finite sums of the form

$$\sum r_i a_i + \sum n_j a_j \qquad (a \in \mathfrak{M},\ r_i \in \mathfrak{o},\ n_j \text{ are integers}).$$

RESIDUE CLASSES. An ideal $\mathfrak{m}$ in $\mathfrak{o}$, being a subgroup of the additive group, defines a partition of $\mathfrak{o}$ into cosets or *residue classes* modulo $\mathfrak{m}$. Two elements a, b are called *congruent modulo* $\mathfrak{m}$, if they belong to the same residue class, i.e., if $a - b \in \mathfrak{m}$. In symbols:

$$a \equiv b \pmod{\mathfrak{m}}$$

or briefly

$$a \equiv b\,(\mathfrak{m}).$$

For "a incongruent to b" we write $a \not\equiv b$.

If, in particular, $\mathfrak{m}$ is a principal ideal (m), we should write $a \equiv b\,((m))$ instead of $a \equiv b\,(\mathfrak{m})$. However, in this case we omit one pair of parantheses and simply write $a \equiv b\,(m)$.

In this way we may arrive, for example, at the ordinary congruence modulo an integer $a \equiv b\,(n)$ (in words: a is congruent to b modulo n) means $a - b$ belongs to (n), i.e. $a - b$ is a multiple of n.

CALCULATIONS WITH CONGRUENCES. The validity of a congruence $a \equiv b$ modulo an ideal $\mathfrak{m}$ is obviously preserved when the same element c is added to both sides, or when both sides are multiplied by c. From this we infer: If $a \equiv a'$ and $b \equiv b'$, then

$$a + b \equiv a + b' \equiv a' + b',$$
$$ab \equiv ab' \equiv a'b';$$

Thus congruences may be added to and multiplied by one another.

We may likewise multiple both sides of a congruence by an ordinary integer n. By taking $n = -1$ we infer that congruences may be substracted from each other.

Thus, we may perform the same operations with congruences as we do with equations; however, in general, we are not permitted to divide; e.g., in the domain of integers we have

$$15 \equiv 3\,(6);$$

but although $3 \not\equiv 0\,(6)$, we cannot infer that $5 \equiv 1\,(6)$.

EXERCISES. 1. Show that in the ring of integers the residue classes modulo an ideal (m) $(m > 0)$ may be represented by the numbers $0, 1, \ldots, m-1$ and may thus be denoted by $\mathfrak{R}_0, \mathfrak{R}_1, \ldots, \mathfrak{R}_{m-1}$.

2. What ideal is generated by the numbers 10 and 13 together in the ring of integers?

3. What does $a \equiv b\ (0)$ mean?

4. All multiples ra of an element a form an ideal $\mathfrak{o}a$. Considering the ring of the even integers, make clear to yourself that this ideal is not necessarily identical with the principal ideal (a).

5. Define also for non-commutative rings the right, left, and two-sided ideal generated by an arbitrary set.

6. What operations on congruences are permissible for non-commutative rings?

Ideals bear the same relation to ring homomorphism as do normal subgroups to group homomorphism. Let us start from the notion of homomorphism.

A homomorphism $\mathfrak{o} \sim \mathfrak{o}$ between two rings defines a partition of the ring $\mathfrak{o}$: a class $\mathfrak{R}_a$ is formed by all elements a having the same image $\bar{a}$. We can characterize these classes more precisely:

The class $\mathfrak{n}$ *of* $\mathfrak{o}$ *to which corresponds the zero element under the homomorphism* $\mathfrak{o} \sim \bar{\mathfrak{o}}$ *is an ideal in* $\mathfrak{o}$, *and the other classes are the residue classes of this ideal.*

PROOF: In the first place $\mathfrak{n}$ is a module; for if a and b become zero under the homomorphism, so does $-b$, and hence also the difference $a-b$; thus, if a and b belong to the class $\mathfrak{n}$, so does $a-b$.

$\mathfrak{n}$ is an ideal; for if a is mapped into zero, and if r is arbitrary, ra is mapped into $\bar{r}\cdot 0 = 0$, thus belonging to $\mathfrak{n}$ again. (In the non-commutative case $\mathfrak{n}$ is even a two-sided ideal.)

The elements $a + c\ (c \in \mathfrak{n})$ of a residue class modulo $\mathfrak{n}$ with a as its representative are carried into $\bar{a} + 0$, hence into $\bar{a}$; therefore, all of them belong to a class $\mathfrak{R}_a$. If, conversely, an element b is carried into $\bar{a}$, then $b-a$ becomes $\bar{a}-\bar{a}=0$; hence $b-a \in \mathfrak{n}$, and b lies in the same residue class as a. This completes the proof.

Thus, to every homomorphism belongs an ideal.

And now for the converse: We start from an ideal $\mathfrak{m}$ in $\mathfrak{o}$ and ask *whether there exists a ring* $\bar{\mathfrak{o}}$ *homomorphic with* $\mathfrak{o}$, *so that the elements of* $\bar{\mathfrak{o}}$ *correspond exactly to the residue classes modulo* $\mathfrak{m}$.

In order to construct such a ring, we proceed in the same manner as we did in Chapter II, Section 10.: As elements of the ring to be constructed we simply choose the residue classes mudulo $\mathfrak{m}$, denote the residue classes $a + \mathfrak{m}$ by $\bar{a}$, and try to define for them an addition and multiplication so that the correspondence $a \to \bar{a}$ constitutes a homomorphism. Thus, for any two residue classes $\bar{a}$, $\bar{b}$ we have to determine a sum class $\bar{a} + \bar{b}$ and a product class $\bar{a} \cdot \bar{b}$ so that all the sums

of the elements of $\bar{a}$ and those of $\bar{b}$ are in the sum class and all products in the product class.

Thus let a be an element in $\bar{a}$, and b one in $\bar{b}$. We tentatively define $\bar{a} + \bar{b}$ as *the class containing* $a + b$, and $\bar{a} \cdot \bar{b}$ as *the class containing* $a \cdot b$. If $a' \equiv a$ is some other element of $\bar{a}$, and $b' \equiv b$ one of $\bar{b}$, then, according to the above[10]

$$a' + b' \equiv a + b,$$
$$a' \cdot b' \equiv a \cdot b;$$

hence $a' + b'$ is in the same residue class as $a + b$, and $a' \cdot b'$ in the same as $a \cdot b$. Our definition of the sum class and product class is therefore independent of the choice of the elements a, b within $\bar{a}$, $\bar{b}$. Now, these classes $\bar{a} + \bar{b}$, $\bar{a} \cdot \bar{b}$ have the desired property, namely that every sum $a' + b'$ is in the sum class $\bar{a} + \bar{b}$, and that every product $a' \cdot b'$ is in the product class of $\bar{a} \cdot \bar{b}$.

To every element a there corresponds a residue class $\bar{a}$; this correspondence is homomorphic, since the sum $a + b$ corresponds to the sum $\bar{a} + \bar{b}$, and the product $\bar{a}\,\bar{b}$ likewise to ab. Hence the residue classes form a ring (Section 12). We shall call this ring the *residue class ring* $\mathfrak{o}/\mathfrak{m}$ of $\mathfrak{o}$ modulo $\mathfrak{m}$. The ring $\mathfrak{o}$ is homomorphically mapped upon $\mathfrak{o}/\mathfrak{m}$ by means of the correspondence mentioned before. For this homomorphism the ideal $\mathfrak{m}$ plays the same rôle as $\mathfrak{n}$ above; it is indeed identical with the set of all elements whose residue class is the null-class.

We note the fundamental importance of the ideals: they enable us to construct rings homomorphic with a given ring. The elements of such a new ring are the residue classes modulo an ideal: with every element a there is associated a residue class $\bar{a}$. Two residue classes are added or multiplied by adding or multiplying any two representatives of these residue classes. $a \equiv b$ implies $\bar{a} = \bar{b}$; *thus the transition to the residue class ring transforms congruences into equalities.*

The special rings homomorphic with $\mathfrak{o}$, as here constructed, i.e. the residue class rings $\mathfrak{o}/\mathfrak{m}$ exhaust virtually all rings homomorphic with $\mathfrak{o}$. For if $\bar{\mathfrak{o}}$ is an arbitrary homomorphic image of $\mathfrak{o}$ there is, as we have seen, a biunique correspondence between the elements of $\bar{\mathfrak{o}}$ and the residue classes of $\mathfrak{o}$ modulo an ideal $\mathfrak{n}$ in $\mathfrak{o}$. The element $\bar{a}$ in $\bar{\mathfrak{o}}$ corresponds to the residue class $\mathfrak{R}_a$. The sum and the product of two residue classes $\mathfrak{R}_a$, $\mathfrak{R}_b$ are given by $\mathfrak{R}_{a+b}$ and $\mathfrak{R}_{ab}$, respectively, and to the latter correspond respectively, the elements

$$\overline{a + b} = \bar{a} + \bar{b}$$

and

$$\overline{a\,b} = \bar{a}\,\bar{b}.$$

Hence the correspondence of the residue classes to the elements of $\bar{\mathfrak{o}}$ constitutes an isomorphism. Thus we have proved the following:

---

[10] All congruences are of course modulo

Any ring $\bar{\mathfrak{o}}$ *homomorphic with* $\mathfrak{o}$ *is isomorphic with a residue class ring* $\mathfrak{o}/\mathfrak{n}$, *where* $\mathfrak{n}$ *is the ideal of those elements whose image in* $\bar{\mathfrak{o}}$ *is the zero. Conversely, every residue class ring* $\mathfrak{o}/\mathfrak{n}$ *is a homomorphic image of* $\mathfrak{o}$. (*Law of homomorphism for rings*).

EXAMPLES FOR THE RESIDUE CLASS RING. In the ring of integers we may denote the residue classes modulo a positive number m by $\mathfrak{R}_0, \mathfrak{R}_1, \ldots, \mathfrak{R}_{m-1}$ (cf. Ex. 1), where $\mathfrak{R}_a$ consists of those numbers which, upon division by m, leave a reminder a. In order to add or multiply two residue classes $\mathfrak{R}_a$, $\mathfrak{R}_b$, we add or multiply their representatives a, b, and reduce the result to its least non-negative remainder modulo m.

EXERCISES. 7. The residue class ring $\mathfrak{o}/\mathfrak{m}$ may have divisors of zero, even though $\mathfrak{o}$ does not have any. Give examples in the ring of integers!

8. The homomorphism $\mathfrak{o} \sim \bar{\mathfrak{o}}$ is a 1-isomorphism, only if $\mathfrak{n} = (0)$.

9. In a field there are no ideals except for the null ideal and the unit ideal. Furnish the proof! What does this imply for the possible homomorphic mappings of a field?

10. In non-commutative rings a homomorphic mapping always defines a *two-sided* ideal, and every two-sided ideal possesses a residue class ring.

11. The ring of the Gaussian integers $a + bi$ (Section 14, Example 1) is isomorphic with the residue class ring modulo *the ideal* $(x^2 + 1)$ in the domain of polynomials in x with integers as coefficients.

17. DIVISIBILITY. PRIME IDEALS

Let $\mathfrak{b}$ be an ideal (or more generally, a module) in the ring $\mathfrak{o}$. If a is an element of $\mathfrak{b}$, we may write $a \equiv 0(\mathfrak{b})$, and we say that *a is divisible by the ideal* $\mathfrak{b}$. If all the elements of an ideal (or of a module) $\mathfrak{a}$ are divisible by $\mathfrak{b}$, then $\mathfrak{a}$ is called *divisible by* $\mathfrak{b}$; this simply means that $\mathfrak{a}$ is a subset of $\mathfrak{b}$. In symbols:

$$\mathfrak{a} \equiv 0(\mathfrak{b}).$$

We call $\mathfrak{b}$ a *divisor* of $\mathfrak{a}$, and $\mathfrak{a}$ a multiple of $\mathfrak{b}$. Thus we see that divide $=$ include, multiple $=$ subset. If, moreover, $\mathfrak{a} \neq \mathfrak{b}$, i.e. if $\mathfrak{a} \subset \mathfrak{b}$, then $\mathfrak{b}$ is called a *proper divisor* of $\mathfrak{a}$, and $\mathfrak{a}$ a *proper multiple* of $\mathfrak{b}$.

For principal ideals in commutative rings with identity $(a) \equiv 0((b))$ simply means $a = rb$, and the concept of divisibility as defined by ideals becomes identical with the ordinary concept.

From now on all rings under consideration will again be commutative.

An ideal $\mathfrak{p}$ in $\mathfrak{o}$ is called a *prime ideal* if its residue class ring $\mathfrak{o}/\mathfrak{p}$ is an integral domain, i.e. one which has no divisors of zero.

If we denote residue classes modulo $\mathfrak{p}$ by bars, as we did before, this condition means that

$\bar{a}\bar{b} = 0$ and $\bar{a} \neq 0$ implies $\bar{b} = 0$.

or, what amounts to the same thing,

$$ab \equiv 0(\mathfrak{p})$$

and

$$a \not\equiv 0(\mathfrak{p})$$

implies

$$b \equiv 0(\mathfrak{p}).$$

for arbitrary a and b in $\mathfrak{o}$. In words: *A product shall be divisible by the ideal $\mathfrak{p}$, only if one factor is divisible by it.*

It is clear that *the unit ideal is always prime;* for the condition $a \not\equiv 0(\mathfrak{o})$ can never be satisfied. *The null ideal is a prime ideal if, and only if, the ring $\mathfrak{o}$ itself is an integral domain.* Further examples of prime ideals are the principal ideals generated by the prime numbers in the ring C of integers, as we shall see later.

An ideal in $\mathfrak{o}$ is called *maximal* if it is not included in any other ideal in $\mathfrak{o}$ except in $\mathfrak{o}$ itself, or in other words, if it has *no proper divisors except the unit ideal* $\mathfrak{o}$. For example, the prime principal ideals (p) in C just mentioned are maximal.

Let $\mathfrak{o}$ be a ring with identity element. *Any maximal ideal $\mathfrak{p}$ in $\mathfrak{o}$, different from $\mathfrak{o}$ itself, is a prime ideal, and the residue class ring $\mathfrak{o}/\mathfrak{p}$ is a field. If, conversely, $\mathfrak{o}/\mathfrak{p}$ is a field, then $\mathfrak{p}$ is maximal.*

PROOF. Let us solve the equation $\bar{x}\bar{a} = \bar{b}$ in the residue class ring for $\bar{a} \neq 0$, supposing $a \not\equiv 0(\mathfrak{p})$. $\mathfrak{p}$ and a together generate an ideal which is a divisor of $\mathfrak{p}$ and (since it contains a) even a proper divisor of $\mathfrak{p}$; hence it must be equal to $\mathfrak{o}$. Therefore, any arbitrary element b of $\mathfrak{o}$ may be written in the form

$$b = \rho + ra \qquad (\rho \in \mathfrak{p},\ r \in \mathfrak{o}),$$

By virtue of the homomorphism of $\mathfrak{o}$ with the residue class ring it follows that

$$\bar{b} = \bar{r}\bar{a},$$

which solves the equation $\bar{x}\bar{a} = \bar{b}$.

Thus it is seen that the residue class ring is a field. Since a field does not have any zero divisors, the ideal $\mathfrak{p}$ is a prime ideal.

If, conversely, $\mathfrak{o}/\mathfrak{p}$ is a field, $\mathfrak{a}$ a proper divisor of $\mathfrak{p}$, a an element of $\mathfrak{a}$ not belonging to $\mathfrak{p}$, then the congruence

$$ax \equiv b(\mathfrak{p})$$

is solvable for every b in $\mathfrak{o}$. It follows that

$$ax \equiv b(\mathfrak{a})$$
$$0 \equiv b(\mathfrak{a});$$

hence $\mathfrak{a} = \mathfrak{o}$, since b may be any element of $\mathfrak{o}$.

Not every prime ideal is maximal. This may be seen from the example of the null ideal in the ring of integers, or less trivially, by the ideal (x) in the integral

polynomial domain $C[x]$ which has among others the ideal $(2, x)$ as a proper divisor. Both (x) and $(2, x)$ are prime ideals, as can be verified easily.

EXERCISES. 1. Prove this last assertion.

2. Discuss the properties of the residue class rings of the ideals (2) and (3) in the ring of integers, and prove that these ideals are prime.

3. Do the same for the ideals (3) and $(1 + i)$ in the ring of the Gaussian integers (Section 14, Example 1). Is the ideal (2) a prime ideal here?

G.C.D. AND L.C.M. The ideal $(\mathfrak{a}, \mathfrak{b})$ generated by the union of two ideals $\mathfrak{a}$ and $\mathfrak{b}$ is also known as the *greatest common divisor* (g.c.d.) of these ideals, since it is a common divisor which is divisible by every common divisor. It is also known as the *sum* of the two ideals, because it evidently consists of all sums $a + b$ where $a \in \mathfrak{a}$, $b \in \mathfrak{b}$.

Similarly, the intersection $\mathfrak{a} \cap \mathfrak{b}$ of two ideals $\mathfrak{a}$, $\mathfrak{b}$ is known as their *least common multiple* (l.c.m.), because it is a common multiple, and because every other multiple is divisible by it.

18. EUCLIDEAN RINGS AND PRINCIPAL IDEAL RINGS

THEOREM. *In the ring C of integers every ideal is a principal ideal.*

PROOF. Let $\mathfrak{a}$ be an ideal in C. If $\mathfrak{a} = (0)$, the proof is completed. If $\mathfrak{a}$ contains a number $c \neq 0$, it also contains $-c$, and one of these numbers is positive. Let a be the least positive number in the ideal $\mathfrak{a}$.

If b is an arbitrary number of the ideal, and if r is the remainder left in the division of b by a, then

$$b = qa + r, \qquad 0 \leqq r < a.$$

Since b and a belong to the ideal, $b - qa = r$ belongs to it as well. Since $r < a$, r must be equal to 0; for a was the least positive number of the ideal. Now we have $b = qa$; i.e. all numbers of the ideal $\mathfrak{a}$ are multiples of a. Hence $\mathfrak{a} = (a)$; therefore $\mathfrak{a}$ is a principal ideal.

Similarly, we prove the following:

If P *is a field, every ideal in the polynomial domain* $\mathsf{P}[x]$ *is a principal ideal.*

We may again assume that $\mathfrak{a} \neq (0)$. Let us choose for a a polynomial of lowest degree in the ideal $\mathfrak{a}$. Since a division algorithm exists also in a polynomial domain, every polynomial b of the ideal can be written as

$$b = qa + r$$

If $r \neq 0$, the degree of r is less than that of a, etc.

An integral domain with identity in which every ideal is principal is called

a *principal ideal ring*. As has just been proved, the ring C of integers, as well as every polynomial ring $\mathsf{P}[x]$, is a principal ideal ring.[11]

In a trivial fashion, every field is furthermore a principal ideal ring. For if an ideal $\mathfrak{a}$ in the field P is not the null ideal, it contains $a^{-1}a = 1$ for an arbitrary $a \neq 0$; hence $\mathfrak{a} = (1)$ is the only ideal besides the null ideal. (Cf. Section 16, Ex. 9.)

The reasoning just applied in two cases may be generalized as follows: Let $\mathfrak{R}$ be a commutative ring in which to every ring element a distinct from zero a non-negative integer $g(a)$ is defined such that

1. for $a \neq 0$ and $b \neq 0$, $ab \neq 0$ and $g(ab) \geqq g(a)$.

2. (Division algorithm) For any two ring elements a, b, where $a \neq 0$, there exists an expression

$$b = qa + r$$

in which either $r = 0$ or $g(r) < g(a)$.

For $R = C$ we have to take $g(a) = |a|$, for $R = \mathsf{P}[x]$ the degree of the polynomial a is $g(a)$. A ring with the properties mentioned is called a *Euclidean ring*. By means of the same reasoning that was previously applied in the two cases $R = C$ and $R = \mathsf{P}[x]$ we can now derive the following theorem:

In a Euclidean ring every ideal is principal, and all elements of the ideal are multiples qa of the generating element a.

Applying this theorem to the unit ideal in particular, i.e. to the entire ring, we see that there exists an a such that all ring elements are multiples qa of it. In particular, a itself is expressible thus:

$$a = ae.$$

For $b = qa$ it follows that

$$qa = qae; \text{ hence } b = be.$$

Thus we have proved the following:

A Euclidean ring always possesses an identity element.

Two elements a, b of a Euclidean ring which are distinct from zero generate an ideal (a, b) consisting of all expressions of the form $ra + sb$. This ideal is principal, i.e. it is generated by an element d. Thus we have:

(1) $$d = ra + sb$$

(2) $$\begin{cases} a = gd \\ b = hd. \end{cases}$$

By (2), d is a common divisor of a and b. By (1), d is also the *greatest common divisor*, i.e. all common divisors of a and b are divisors of d. Thus we may state the following theorem: *In a principal ideal ring any two elements a, b have a greatest common divisor d expressible in the form (1).* As a special case, this is true for the ring of integers and for the ring $\mathsf{P}[x]$:

[11] An elementary investigation of the conditions which an integral domain has to satisfy, in order that every ideal in it be principal, was published by H. Hasse in *Crelles J. f. Math.* Vol. 159, pp. 3-12, 1928.

The greatest common divisor is usually designated by $d = (a, b)$. A more exact notation, however, would be $(d) = (a, b)$; for it is only the ideal (d) and not the element d that is uniquely determined by a and b. If $(a, b) = 1$, then a and b are called *relatively prime.*

The above existence proof for the g.c.d. does not provide a tool for actually computing it. For Euclidean rings we can employ the method of successive divisions explained by Euclid in the 7th book of his Elements (Theorems 1 and 2). This method is known as the *Euclidean algorithm.*

Let two ring elements a_0, a_1 be given, and let $g(a_1) \leqq g(a_0)$. Then, by the division algorithm, we let

$$a_0 = q_1 a_1 + a_2 \qquad g(a_2) < g(a_1)$$
$$a_1 = q_2 a_2 + a_3 \qquad g(a_3) < g(a_2)$$

and continue in this way, until some division yields the remainder 0:

$$a_{s-1} = q_s a_s.$$

Then all numbers $a_0, a_1, a_2, \ldots, a_s$ are of the form $ra_0 + sa_1$. Every divisor of a_s (and, in particular, a_s itself) is, according to the last equation, also a divisor of a_{s-1}, furthermore a divisor of a_{s-2}, and finally of a_1 and a_0. Hence a_s is the greatest common divisor of a_0 and a_1.

These considerations may also be applied to the non-commutative case, provided there exists a left-sided *and* a right-sided division algorithm:

$$b = q_1 a + r_1 = a q_2 + r_2, \qquad g(r_1) < g(a), \qquad g(r_2) < g(a).$$

It follows that every left ideal contains an element a such that all elements of the ideal are left multiples qa of a, and that likewise every right ideal contains an element a such that all elements of the ideal are right multiples aq of a. A two-sided ideal possesses a generating element a such that all elements are both right and left multiples of a. Applying this to the unit ideal, it follows that there exist a left and a right identity, i.e. that there is an identity element.

Finally, we may prove, as above, the existence of a left-sided as well as of a right-sided g.c.d. of two elements a, b.

The most important example of a non-commutative Euclidean ring is the polynomial ring $\mathsf{P}[x]$ over a skew field P.

EXERCISES. 1. The relation $(a, b) = d$ remains valid when the ring $\mathfrak{o}$ is extended to any larger ring $\mathfrak{o}$.

2. The elements ε of a Euclidean ring having an inverse ε^{-1} in the ring are characterized by $g(\varepsilon) = g(1)$.

3. Every element a of order $r \cdot s$ in a group $\mathfrak{G}$ is the product of a uniquely determined element $a^{\lambda r}$ of order s and a uniquely determined element $a^{\mu s}$ of order r, provided the numbers r and s are relatively prime:

$$(r, s) = 1.$$

4. A cyclic group of order n with the generating element a can also be generated by any power a^{μ}, provided $(\mu, n) = 1$.

ANOTHER EXAMPLE OF A EUCLIDEAN RING. Let us consider the ring of the Gaussian integers $a+bi$ (end of Section 14).

The definition of a product in this ring was

$$(a+bi)(c+di)=(ac-bd)+(ad+bc)i\ .$$

Now defining the "norm" of a number $\alpha=a+bi$ by

$$N(\alpha)=(a+bi)(a-bi)=a^2+b^2$$

we easily derive the equation

$$N(\alpha\beta)=N(\alpha)\cdot N(\beta)\,.$$

The norm $N(\alpha)$ is an ordinary integer which, being the sum of two squares, vanishes only when α vanishes, and which is positive in any other case. From (3) it follows that a product $\alpha\beta$ vanishes only when α or β vanishes; hence the ring is an integral domain.

According to Section 13, a quotient field exists. If $\alpha=a+bi\neq 0$, then $\alpha^{-1}=\frac{a-bi}{N(\alpha)}$; thus the numbers of the quotient field may be expressed by $\frac{a}{n}+\frac{b}{n}i$ (a, b, n are integers). These "fractional numbers" form the "Gaussian number field" (Section 14, Example 1). The definition of the norm and equation (3) hold for the elements of this field as well.

In order to arrive at a division algorithm for the ring of the Gaussian integers, we have to find for a given α and $\beta\neq 0$ a number $\alpha-\lambda\beta$ having norm less than β. Let us first determine a fractional number $\lambda'=a'+b'i$ so that $\alpha-\lambda'\beta=0$; then let us replace a' and b' by the nearest integers a and b, and put $\lambda=a+bi$, $\lambda'-\lambda=\varepsilon$. Then we have:

$$\alpha-\lambda\beta=\alpha-\lambda'\beta+\varepsilon\beta=\varepsilon\beta,$$

$$N(\alpha-\lambda\beta)=N(\varepsilon)\,N(\beta),$$

$$N(\varepsilon)=N(\lambda'-\lambda)=(a'-a)^2+(b'-b)^2\leqq(\tfrac{1}{2})^2+(\tfrac{1}{2})^2<1,$$

$$N(\alpha-\lambda\beta)<N(\beta)\,.$$

Thus we have found a "division algorithm," which proves that the ring is a Euclidean ring.

EXERCISE. 5. The ring of the numbers $a+b\varrho$, which, as a linear associative algebra over the ring of integers, is defined by the basis elements 1, ϱ and the operation

$$\varrho^2=-\varrho-1$$

is to be treated in the same manner; similarly, the rings of the numbers $a+b\sqrt{2}$, $a+b\sqrt{-2}$. Why does the method fail for $a+b\sqrt{-3}$ and $a+b\sqrt{-5}$? Is the ideal $(2,\ 1+\sqrt{-3})$ in the first mentioned ring a principal ideal?

BIBLIOGRAPHY. Concerning the question whether the Euclidean algorithm or its generalization exists in arbitrary principal ideal rings, see H. Hasse in *J. reine u. angew. Math.* Vol. 159 (1928), pp. 3-12. Investigations as to the question in what algebraic number rings the Euclidean algorithm is valid were carried out by O. Perron (*Math. Ann.* Vol. 107, p. 489), A. Oppenheim (*Math. Ann.* Vol. 109, p. 349), E. Berg (*Kgl. Fysiogr. Sällskapets Lund Förhandl.* Vol. 5N 5), N. Hofreiter (*Mh. Math. Physik* Vol. 42, p. 397), H. Behrbohm and L. Redei (*J. reine u. angew. Math.* Vol. 174, p. 198).

19. FACTORIZATION

In this section we shall be concerned merely with integral domains containing the identity. Let us first investigate what we mean by prime elements or indecomposable elements in these domains. We shall consider only ring elements distinct from zero, even when it is not expressly stated.

A prime number in the ring of integers may always be decomposed into factors, even in two ways:

$$p = p \cdot 1 = (-p) \cdot (-1).$$

However, one of these factors is always a "unit", i.e. a number ε, whose inverse ε^{-1} is likewise in the ring $+1$ and -1 are units.

If, in general, an integral domain with the identity is given, then by a unit we understand an element ε which possesses an inverse ε^{-1} in the domain. Then ε^{-1}, obviously, is also a unit.

If ε is a unit, then every element a admits a decomposition

$$a = a\varepsilon^{-1} \cdot \varepsilon .$$

Such decompositions where one factor is a unit may be called "trivial decompositions".

An element $p \neq 0$ which admits only trivial decompositions of the kind $p = ab$, where a or b is a unit, is called an *indecomposable element* or a *prime element.* (In case of integers we say: *prime number*; [12] in case of polynomials: *irreducible polynomial.*)

Two quantities, such as a and $b = a\varepsilon^{-1}$, which differ only by a unit as factor, are sometimes called "associates". Either one is a divisor of its associate, and for their respective principal ideals we have:

$$(a) \subseteq (b), \qquad (b) \subseteq (a), \quad \text{hence } (b) = (a).$$

Thus two associates generate the same principal ideal.

If, conversely, either of the two quantities a and b divides the other one, viz.

$$a = bc, \qquad b = ad,$$

it follows that

$$b = bcd, \quad \text{hence} \quad 1 = cd, \qquad c = d^{-1};$$

hence c and d are units, and a and b are associates.

If c is a divisor of a, but not an associate of a, i.e. $a = cd$, and d are not units, then c is called a *proper divisor* of a. In this case a is not a divisor of c, and the ideal (c) is a proper divisor of the ideal (a). For if a were a divisor of c, say $c = ab$, then we would have

$$a = cd = abd$$
$$1 = bd$$

and d would be a unit, contrary to our assumption.

A prime element may also be defined as an element distinct from zero which does not possess any proper divisors except units.

If, in a Euclidean ring, b is a proper divisor of a, then $g(b) < g(a)$.

PROOF: The division of b and a leaves a remainder, viz.

---

[12] By prime numbers we usually understand only the positive prime numbers $\neq 1$, such as 2, 3, 5, 7, 11, . . .

$$b = aq + r, \quad g(r) < g(a).$$

Equating $a = bc$, it follows that

$$r = b - aq = b(1 - cq),$$

$$g(r) \geqq g(b); \text{ hence } g(b) \leqq g(r) < g(a).$$

In a Euclidean ring every element a distinct from zero is a product of prime elements:

$$a = p_1 p_2 \cdots p_r.$$

PROOF: We apply the method of complete induction on $g(a)$: Let the assertion be true for all elements b with $g(b) < n$, and let $g(a) = n$. If a is prime: $a = p$, there is nothing more to be proved. However, if a is decomposable: $a = bc$, where b and c are proper divisors of a, then

$$g(b) < g(a), \qquad g(c) < g(a).$$

By the induction hypothesis, b and c are products of prime elements. Hence $a = bc$ is also a product of prime elements.

We now investigate the uniqueness of factorization into primes $a = p_1 p_2 \cdots p_r$ and consider not only Euclidean rings, but arbitrary principal ideal rings.

In a principal ideal ring an indecomposable element, other than a unit, generate a maximal prime ideal. (By a theorem previously proved, the residue class ring of this ideal is a field.)

PROOF: If p is indecomposable, then it has no proper divisors except units, and therefore (since every ideal is a principal ideal) the ideal (p) has no proper ideal divisors except the unit ideal.

REMARK. The solvability of the equation $\overline{ax} = \overline{b}$ in the residue class ring or of the congruence $ax \equiv b\,(p)$ in the given ring may, of course, readily be seen from the fact that, for $a \not\equiv 0\,(p)$, we have $(a, p) = 1$. For this implies

$$1 = ar + ps,$$
$$b = arb + psb$$
$$b \equiv arb\ (p)$$

On the basis of this remark we may, for particular cases, actually compute the solution of the congruence mentioned by means of the Euclidean algorithm.

We infer at once:

If a product is divisible by the prime element p, so is a factor of the product; for the residue class ring has no divisors of zero.

EXERCISES. 1. Solve the congruence

$$6x \equiv 7\,(19).$$

2. What is the inverse element of the residue class of 6 in the residue class field of integers modulo (19)?

We are now in a position to prove the *theorem of uniqueness of prime factorization in principal ideal rings.* Let

(1) $$a = p_1 p_2 \cdots p_r = q_1 q_2 \cdots q_s$$

be two decompositions of the same number a in a principal ideal ring. We shall exclude the trivial case where a is a unit and where, consequently, all p_i and q_j are units. Then we may assume that p_1 and q_1 are not units, and that all possible units among the factors p_i and q_j are combined with the factor p_1 and q_1, respectively. Thus let the p_i and q_j not be units. Now we state: $r=s$, *and the* p_i and q_j *are identical, except for their order and for unit factors.*

For $r=1$ the proof is clear; for since $a=p_1$ is prime, the product $q_1 \cdots q_s$ can contain only one factor $q_1=p_1$ Thus we may proceed by induction on r. Since p_1 divides the product $q_1 \cdots q_s$, p_1 must divide one of the factors q_t. With the q rearranged, p_1 will divide q_1:

$$q_1=\varepsilon_1 p_1. \tag{2}$$

Here ε_1 must be a unit, or else q_1 would not be prime. On substituting (2) in (1) and dividing by p_1, we obtain

$$p_2 \cdots p_r=(\varepsilon_1 q_2) q_3 \cdots q_s. \tag{3}$$

By the induction hypothesis, the factors on the left and right side of (3) must be the same, except for the unit factors. Since p_1 is identical with q_1, except for the unit factor ε_1, the proof is completed.

From the theorems proved we infer: *The elements of a Euclidean ring are uniquely expressible as products of prime elements, except for units and for the order of the factors.* This is particularly true for the integers, for the polynomials in one variable with coefficients from a field, and for the Gaussian integers.

EXERCISES. 3. The integral polynomials $f(x)$ modulo any prime number p are uniquely decomposable into factors which are indecomposable modulo p.

4. What are the units of the Gaussian number ring? Decompose into prime factors the numbers 2, 3, 5 in this ring.

5. For the number 4 in the ring of the numbers $a+b\sqrt{-3}$ there are two substantially different factorizations into indecomposable factors:

$$4=2\cdot 2=(1+\sqrt{-3})(1-\sqrt{-3}).$$

6. In a principal ideal ring the residue classes modulo a consisting of elements relatively prime to a form a group under multiplication.

In the following chapter we shall see that there are rings other than principal ideal rings for which the unique factorization theorem holds. For all such rings we shall now prove the following theorem:

If in o *every element is uniquely decomposable into prime elements, then every indecomposable element p generates a prime ideal, and every decomposable element distinct from zero generates a non-prime ideal.*

PROOF: Let p be indecomposable. If $ab \equiv 0(p)$, then the factor p must occur, when ab is factored. This factorization, however, is obtained by combining the

factorizations of a and b; therefore, the factor p must already occur in a or b, whence $a \equiv 0(p)$ or $b \equiv 0(p)$.

Now let p be decomposable: $p = ab$, where a and b are proper divisors of p. Then it follows that $ab \equiv 0(p)$, $a \not\equiv 0(p)$, $b \not\equiv 0(p)$. Therefore, the ideal (p) is not prime.

EXERCISES. 7. Prove that for all rings with unique factorization every two or more elements have a "greatest common divisor" and a "least common multiple", both of them being determined except for unit factors.

REMARK. For rings of the kind considered, the g.c.d. in the sense of an element is not always the same as the g.c.d. in the sense of an ideal. For example, in the polynomial domain of a variable x with integer coefficients the elements 2 and x have no common divisors except units; but the ideal $(2, x)$ is not the unit ideal. (In the next chapter it will be proved that there is a unique factorization in this ring.)

CHAPTER IV

POLYNOMIALS

CONTENTS: Simple theorems on polynomials in one and several variables with coefficients in a commutative ring $\mathfrak{o}$ or field Σ.

20. DIFFERENTIATION

In this section we shall define the differential quotients of polynomials for arbitrary polynomial domains $\mathfrak{o}[x]$ without making use of the notion of continuity.

Let $f(x)=\sum a_i x^i$ be a polynomial in $\mathfrak{o}[x]$. If we form in a polynomial domain $\mathfrak{o}[x, h]$ the polynomial $f(x+h)=\sum a_i(x+h)^i$ and develop it in powers of h, we obtain:

$$f(x+h)=f(x)+hf_1(x)+h^2f_2(x)+\cdots$$

or

$$f(x+h)\equiv f(x)+h\cdot f_1(x)\,(\mathrm{mod}\,h^2).$$

The (uniquely determined) coefficient $f_1(x)$ of the first power of h is called the *derivative* of $f(x)$ and is always denoted by $f'(x)$ . . $f'(x)$ may also be obtained by forming the difference $f(x+h)-f(x)$, by dividing it by the rational integral factor h contained therein, and by setting $h=0$ in the polynomial thus obtained. From this it is readily seen that the definition of the derivative is identical with the conventional definition of the *differential quotient* $\lim\limits_{h\to 0}\frac{f(x+h)-f(x)}{h}$ if $\mathfrak{o}$ is e.g. the field of real numbers. Therefore, the derivative may also be denoted by $\frac{df}{dx}$ or $\frac{d}{dx}f(x)$, or, if f has other variables besides x, by $\frac{\partial f}{\partial x}$.

The following rules for differentiation hold:

(1) $$(f+g)'=f'+g',$$

(2) $$(fg)'=f'g+fg'.$$

PROOF (1):

$$f(x+h)+g(x+h)\equiv f(x)+hf'(x)+g(x)+hg'(x)\,(\mathrm{mod}\,h^2).$$

PROOF (2):

$$f(x+h)g(x+h) \equiv \{f(x)+hf'(x)\}\{g(x)+hg'(x)\}$$
$$\equiv f(x)g(x)+h\{f'(x)g(x)+f(x)g'(x)\} \pmod{h^2}.$$

Similiarly, we may prove more generally:

$$(f_1+\cdots+f_n)' = f_1'+\cdots+f_n', \tag{3}$$

$$(f_1 f_2 \cdots f_n)' = f_1' f_2 \cdots f_n + f_1 f_2' \cdots f_n + \cdots + f_1 f_2 \cdots f_n'. \tag{4}$$

From (4) follows

$$(a x^n)' = n a x^{n-1}. \tag{5}$$

From (3) and (5) follows

$$\left(\sum_0^n a_k x^k\right)' = \sum_0^n k a_k x^{k-1}.$$

This formula may serve as an alternate definition of the differential quotient.

EXERCISES. 1. Let $F(z_1, \ldots, z_m)$ be a polynomial, and let $F_\nu = \frac{\partial F}{\partial z_\nu}$. Prove the formula

$$\frac{d}{dx} F(f_1(x), \ldots, f_m(x)) = \sum_1^m F_\nu(f_1, \ldots, f_m) \frac{df_\nu}{dx}.$$

2. Derive "Euler's differential equation" for homogeneous polynomials of degree r,

$$\sum_\nu \frac{\partial f}{\partial x_\nu} x_\nu = r f$$

from the equation

$$f(h x_1, \ldots, h x_n) = h^r f(x_1, \ldots, x_n).$$

3. Give an algebraic definition for the derivative of a rational fractional function $\frac{f(x)}{g(x)}$ with coefficients in a field, and prove the well known rules for differentiation of sums, products, and quotients.

21. THE ZEROS OF A POLYNOMIAL

Let $\mathfrak{o}$ be an integral domain containing the identity.

An element α of $\mathfrak{o}$ is called a *zero* or a *root* of a polynomial $f(x)$ *in* $\mathfrak{o}[x]$ if $f(\alpha)=0$. The following theorem holds:

If α is a root of $f(x)$, then $f(x)$ is divisible by $x-\alpha$.

PROOF: Dividing $f(x)$ by $x-\alpha$, we obtain

$$f(x) = q(x)\cdot(x-\alpha)+r,$$

where r is a constant. Substituting $x=\alpha$, we get

$$0 = r;$$

hence

$$f(x) = q(x)\cdot(x-\alpha). \qquad \text{Q.E.D.}$$

If $\alpha_1, \ldots, \alpha_k$ are different roots of $f(x)$, then $f(x)$ is divisible by the product $(x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_k)$.

PROOF: For $k=1$ the theorem has just been proved. If the theorem is proved for the value $k-1$, we have

$$f(x)=(x-\alpha_1)\cdots(x-\alpha_{k-1})g(x).$$

Substituting $x=\alpha_k$, we obtain:

$$0=(\alpha_k-\alpha_1)\cdots(\alpha_k-\alpha_{k-1})g(\alpha_k).$$

Since $\mathfrak{o}$ has no divisors of zero, and since $\alpha_k\neq\alpha_1,\ldots,\alpha_k\neq\alpha_{k-1}$, this implies

$$g(\alpha_k)=0,$$

and hence, by the previous theorem,

$$g(x)=(x-\alpha_k)\cdot h(x),$$

$$f(x)=(x-\alpha_1)\cdots(x-\alpha_{k-1})(x-\alpha_k)h(x). \qquad \text{Q.E.D.}$$

COROLLARY: *An n-th degree polynomial distinct from zero has in any integral domain at most n roots.*

This corollary holds also in integral domains without an identity, since such a domain may always be embedded in a field (with identity). However, it does not hold for rings with divisors of zero; e.g., in the residue class ring modulo 16 the polynomial x^2 has the roots 0, 4, 8, 12, and there are even rings in which the same polynomial has an infinite number of roots (Section 11, Ex. 3). For non-commutative rings the corollary does not hold either; for in the field of quaternions (Section 14, Ex. 2) the polynomial x^2+1 has the roots $\pm i$, $\pm j$, $\pm k$ (and infinitely many more).

If $f(x)$ *is divisible by* $(x-\alpha)^k$, but not by $(x-\alpha)^{k+1}$, then α is called a *root of multiplicity* k of $f(x)$. Now the following theorem holds:

A root of multiplicity k of $f(x)$ is at least a root of multiplicity $k-1$ of the derivative $f'(x)$.

PROOF: From $f(x)=(x-\alpha)^k g(x)$ we find

$$f'(x)=k(x-\alpha)^{k-1}g(x)+(x-\alpha)^k g'(x);$$

hence $f'(x)$ is divisible by $(x-\alpha)^{k-1}$.

Similarly, we may prove: *A root of $f(x)$ of multiplicity* 1 *(simple root) is not at the same time a root of the derivative $f'(x)$.*

We now proceed to prove some theorems on the roots of polynomials in several variables.

If a polynomial $f(x_1,\ldots,x_n)$ is distinct from zero, and if we make available to each of the indeterminates $x_1,\ldots,x_n$ an infinite set of special values in $\mathfrak{o}$ or in an integral domain including $\mathfrak{o}$, then there exists at least one system of values $x_1=\alpha_1,\ldots,x_n=\alpha_n$, for which $f(\alpha_1,\ldots,\alpha_n)\neq 0$.

PROOF: $f(x_1,\ldots,x_n)$ considered as a polynomial in x_n with coefficients in the integral domain $\mathfrak{o}[x_1,\ldots,x_{n-1}]$ has at most a finite number of roots; hence there exists in the infinite set of values available for x_n a value α_n such that

$$f(x_1,\ldots,x_{n-1},\alpha_n)\neq 0.$$

This expression may now be treated as a polynomial in x_{n-1}; thus we obtain a value α_{n-1}, for which

$$f(x_1, x_1, \ldots, x_{n-2}, \alpha_{n-1}, \alpha_n) \neq 0,$$

etc.

COROLLARY: *If for all special values x_i in an infinite integral domain the polynomial $f(x_1, \ldots, x_n)$ becomes zero, it vanishes ("identically").*

We should bear in mind that in algebra the vanishing of a polynomial in $x_1, \ldots, x_n$ means the vanishing of all coefficients, but it is not defined as the vanishing of the value of the polynomial for all values which may be substituted for $x_1, \ldots, x_n$. Thus, the corollary just established does not constitute a tautology. For finite integral domains [1] and for many rings with divisors of zero the theorem does not hold.

EXERCISE. Extend the above corollary to a finite set of polynomials $f_i(x_1, \ldots, x_n)$, no one of which vanishes identically.

22. INTERPOLATION FORMULAE

Let us return to the polynomials in one variable, but let us assume that the domain of coefficients is a *field*. According to the theorems proved, two polynomials of degree $\leqq n$, whose values coincide at $n+1$ points are equal; for their difference has $n+1$ roots and is at most of degree n. Thus, there is at most one polynomial which, at $n+1$ different points $\alpha_0, \ldots, \alpha_n$ assumes given values $f(\alpha_i)$. Now, there is always one polynomial of degree $\leqq n$, which assumes the given values at these points; it is the polynomial

$$(1) \qquad f(x) = \sum_{i=0}^{n} \frac{f(\alpha_i)(x-\alpha_0)\cdots(x-\alpha_{i-1})(x-\alpha_{i+1})\cdots(x-\alpha_n)}{(\alpha_i-\alpha_0)\cdots(\alpha_i-\alpha_{i-1})(\alpha_i-\alpha_{i+1})\cdots(\alpha_i-\alpha_n)}.$$

Thus there exists one, and only one, polynomial of degree $\leqq n$, which, at $n+1$ points α_i assumes given values $f(\alpha_i)$. This polynomial is given by formula (1). This formula (1) is known as *Lagrange's interpolation formula*, because it permits us to compute the values of a rational integral function of degree n for all intermediate argument values, once its values are known for $n+1$ argument values.

A polynomial having the desired properties may also be obtained by means of *Newton's interpolation formula*

$$(2) \qquad \begin{aligned} f(x) = {} & \lambda_0 + \lambda_1(x-\alpha_0) + \lambda_2(x-\alpha_0)(x-\alpha_1) + \cdots \\ & + \lambda_n(x-\alpha_0)(x-\alpha_1)\cdots(x-\alpha_{n-1}), \end{aligned}$$

[1] Example: The polynomial $x^2 + x$ vanishes for all x in the field $C/(2)$ without vanishing itself.

The coefficients $\lambda_0, \ldots, \lambda_n$ may be successively determined by substituting the values $x = \alpha_0, \ldots, x = \alpha_n$. Doing this, we get for every λ_i a linear equation in which the coefficient of this λ_i has the value

$$(\alpha_i - \alpha_0)(\alpha_i - \alpha_1) \cdots (\alpha_i - \alpha_{i-1}) \neq 0$$

and in which the other terms contain only λ with smaller subscripts.

The computation is preferably carried out as follows:

First substitute $x = \alpha_0$ in (2) which gives

$$f(\alpha_0) = \lambda_0.$$

Subtracting this from (2) and dividing by $x - \alpha_0$, we obtain

$$\frac{f(x) - f(\alpha_0)}{x - \alpha_0} = \lambda_1 + \lambda_2(x - \alpha_1) + \cdots + \lambda_n(x - \alpha_1) \cdots (x - \alpha_{n-1}). \tag{3}$$

We call the left member $f(\alpha_0, x)$. Substituting $x = \alpha_1$ in (3), we have

$$f(\alpha_0, \alpha_1) = \lambda_1.$$

We subtract this from (3), divide by $x - \alpha_1$, and obtain

$$\frac{f(\alpha_0 x) - f(\alpha_0, \alpha_1)}{x - \alpha_1} = \lambda_2 + \lambda_3(x - \alpha_2) + \cdots + \lambda_n(x - \alpha_2) \cdots (x - \alpha_{n-1}).$$

We call the left hand member $f(\alpha_0, \alpha_1, x)$. Putting $x = \alpha_2$, it follows that

$$f(\alpha_0, \alpha_1, \alpha_2) = \lambda_2.$$

We may proceed in this manner. We define by complete induction:

$$f(\alpha_0, \ldots, \alpha_k, x) = \frac{f(\alpha_0, \ldots, \alpha_{k-1}, x) - f(\alpha_0, \ldots, \alpha_{k-1}, \alpha_k)}{x - \alpha_k} \tag{4}$$

and we find, as above

$$f(\alpha_0, \ldots, \alpha_{k-1}, x) = \lambda_k + \lambda_{k+1}(x - \alpha_k) + \cdots + \lambda_n(x - \alpha_k) \cdots (x - \alpha_{n-1}),$$

$$f(\alpha_0, \ldots, \alpha_k) = \lambda_k. \tag{5}$$

We call $f(\alpha_0, \ldots \alpha_k)$ the k-th difference quotient of the function $f(x)$ at the points $\alpha_0, \ldots, \alpha_k$. By (4) we have:

$$\left\{\begin{array}{l} f(\alpha_0, \alpha_1) = \dfrac{f(\alpha_1) - f(\alpha_0)}{\alpha_1 - \alpha_0} \\ f(\alpha_0, \alpha_1, \alpha_2) = \dfrac{f(\alpha_0, \alpha_2) - f(\alpha_0, \alpha_1)}{\alpha_2 - \alpha_1} \\ f(\alpha_0, \ldots, \alpha_r) = \dfrac{f(\alpha_0, \ldots, \alpha_{n-2}, \alpha_n) - f(\alpha_0, \ldots, \alpha_{n-2}, \alpha_{n-1})}{\alpha_n - \alpha_{n-1}}. \end{array}\right. \tag{6}$$

The k-th difference quotient may also be defined as the coefficient of x^k in that polynomial $\varphi_k(x)$ of degree $\leqq k$, which takes the values $f(\alpha_0), \ldots, f(\alpha_k)$ at the points $\alpha_0, \ldots, \alpha_k$. For, by Newton's interpolation formula, this polynomial is given by

$$\varphi_k(x) = \lambda_0 + \lambda_1(x - \alpha_0) + \cdots + \lambda_k(x - \alpha_0) \cdots (x - \alpha_{k-1}),$$

and the coefficient of x^k in this expression is exactly $\lambda_k = f(\alpha_0, \ldots, \alpha_k)$.

It follows from the last mentioned definition that the k-th difference quotient is independent of the order (i.e. of the labeling) of the points $\alpha_0, \ldots, \alpha_k$. In practice, this property is utilized, for example, if $\alpha_0, \ldots, \alpha_n$ are given as rational numbers

in natural order, by forming the difference quotients always for successive points α_ν only, and by using instead of (6) the formulae

$$f(\alpha_0, \alpha_1, \ldots, \alpha_k) = \frac{f(\alpha_1, \ldots, \alpha_k) - f(\alpha_0, \ldots, \alpha_{k-1})}{\alpha_k - \alpha_0} \tag{7}$$

which are obtained by interchanging the α_ν in (6). The difference quotients may then be arranged as in the following array:

$$\begin{array}{llll} f(\alpha_0) & & & \\ & f(\alpha_0, \alpha_1) & & \\ f(\alpha_1) & & f(\alpha_0, \alpha_1, \alpha_2) & \\ & f(\alpha_1, \alpha_2) & & \ldots \\ f(\alpha_2) & & f(\alpha_1, \alpha_2, \alpha_3) & \\ & f(\alpha_2, \alpha_3) & \ldots & \\ f(\alpha_3) & \ldots & & \\ \ldots & & & \end{array}$$

By (7), every succeeding column is obtained by forming the first difference quotients from the preceding column. The array may be continued downward at pleasure by using more and more argument values. If $f(x)$ is a polynomial of the n-th degree, we obtain a constant in the n-th column everywhere, namely the coefficient λ_n of x^n, and in the $(n+1)$-st column we would constantly get zero.

Arithmetic series of higher order. We assume that the underlying field includes the field of rational numbers, and that the points $\alpha_0, \alpha_1, \alpha_2, \ldots$ are chosen as successive integers, say 0, 1, 2, . . . If we form the above array of difference quotients, the denominators $\alpha_k - \alpha_0$, $\alpha_{k+1} - \alpha_1, \ldots$ which, according to (7), appear when the difference quotients of the $(k+1)$-st column are computed, are all equal to k. Multiplying the second column by 1, the third by 2, the fourth by $2 \cdot 3$, and in general the $(k+1)$-st by k, we obtain, instead of the array of difference quotients, *the array of differences*

$$\left|\begin{array}{llll} a_0 & & & \\ & \Delta a_0 & & \\ a_1 & & \Delta^2 a_0 & \\ & \Delta a_1 & & \ldots \\ a_2 & & \Delta^2 a_1 & \\ & \Delta a_2 & \vdots & \\ a_3 & \vdots & & \\ \vdots & & & \end{array}\right. \tag{8}$$

In this procedure we put $f(\alpha_\nu) = a_\nu$. Δa_ν stands for $a_{\nu+1} - a_\nu$; $\Delta^2 a_\nu$ stands for $\Delta \Delta a_\nu = \Delta a_{\nu+1} - \Delta a_\nu$, etc. If $a_0, a_1, \ldots$ are the values of a polynomial of the n-th degree, then, according to what was stated above, the n-th differences are constant and the $(n+1)$-st differences are zero. The polynomial itself is given by (2) with

$$\lambda_k = \frac{\Delta^k a_0}{k!}. \tag{9}$$

The converse theorem also holds:

If the $(n+1)$*-st differences of the sequence* $a_0, a_1, a_2, \ldots$ *are zero, then* $a_0, a_1, \ldots$ *are the values of a polynomial* $f(x)$ *of degree n given by* (2) *and* (9).

For if we form the array of differences, starting with the values of the polynomial $f(x)$ and compare it with the given array (8), we see that the initial elements $a_0, \Delta a_0, \Delta^2 a_0, \ldots, \Delta^n a_0$ of the columns agree in the two arrays, while the $(n+1)$-st column contains only zeros in either array. Hence it follows in order that the elements of the n-th column, the elements of the $(n-1)$-st column, . . . , and finally those of the first column are all identical for the two arrays.

Starting with the last column, we may, by the same method, compute all elements of the array (8), if the initial elements $\Delta^k a_0 = k!\,\lambda_k (k = 0,1,\ldots,n)$ of the columns are given. The following example ($n=3$, $a_0=0$, $\Delta a_0=1$, $\Delta^2 a_0=6$, $\Delta^3 a_0=6$) will explain the computation:

$$\begin{array}{ccccc}
0 & & & & \lambda_0 = 0 \\
& 1 & & & \\
1 & & 6 & & \lambda_1 = 1 \\
& 7 & & 6 & \\
8 & & 12 & & \lambda_2 = \tfrac{6}{2} = 3 \\
& 19 & & 6 & \\
27 & & 18 & & \lambda_3 = \tfrac{6}{6} = 1 \\
& 37 & & 6 & \\
64 & & 24 & & \\
& 61 & & & \\
125 & & & &
\end{array}$$

$$f(x) = \lambda_0 + \lambda_1 x + \lambda_2 x(x-1) + \lambda_3 x(x-1)(x-2) =$$
$$= x + 3x(x-1) + x(x-1)(x-2) = x^3.$$

By an arithmetic series of zero-th order we shall mean a sequence of identical numbers $c, c, c, \ldots$, and by an arithmetic series of n-th order a sequence of numbers such that its sequence of differences is an arithmetic series of $(n-1)$-st order. Then it is obvious that the column of the array (8) forms an arithmetic series of the n-th order, provided the $(n+2)$-nd column consists of zeros only. Consequently, what was proved above may be formulated thus:

The values of a polynomial $f(x)$ *of degree n at the points* 0, 1, 2, 3, . . . *form an arithmetic series of the n-th order, and every arithmetic series of the n-th order consists of the values of a polynomial of at most degree n at those points.* The polynomial $f(x)$ itself is obtained from (2) and (9). Thus, the generic term a_x of an arithmetic series of order n is given by the formula

$$a_x = f(x) = a_0 + (\Delta a_0)\,x + \frac{\Delta^2 a_0}{2}\,x(x-1) + \cdots + \frac{\Delta^n a_0}{n!}\,x(x-1)\cdots(x-n+1).$$

A practical application of the array of differences (8) can be found in the in-

terpolation and integration of functions given by numerical tables (e.g. by tables obtained empirically). If $a_0, a_1, a_2, \ldots$ are the values of a function $\varphi(x)$ for equidistant argument values $\alpha_0, \alpha_0+h, \alpha_0+2h, \ldots$, it will be seen that, for functions with a regular course and for not too great an interval h, the second, third, fourth, or in the worst case the fifth differences become practically zero, which shows that in some adjacent intervals the function behaves almost exactly like the polynomial of at most degree four. Thus, for numerical interpolation or integration, the function may be replaced by the polynomial which assumes the table values at 2 to 5 successive points. Interpolation is carried out by means of formula (2). In most cases linear or quadratic interpolation is sufficient, which means that only the first and second differences are needed, and the higher ones may be neglected. When differences $\Delta^k a_\nu$ are converted into difference quotients, powers of the length of the interval h appear besides the factors $k!$; accordingly, instead of (9), we have to use the formula

$$\lambda_k = \frac{\Delta^k a_0}{k!\, h^k} .$$

For argument values $\alpha_0, \alpha_1, \ldots$ no longer equidistant we have to form difference quotients (7) right at the outset instead of the differences $\Delta^k a_\nu$. Further details of the computation as well as error estimates will be found in special text books.[2]

EXERCISES. 1. The partial sums $s_m = \sum_{\nu=0}^{m-1} a_\nu$ of an arithmetic series of the n-th order (where $s_0 = 0$) form an arithmetic series of the $(n+1)$-st order. Derive from this the formula for the sum

$$s_m = m a_0 + \binom{m}{2} \Delta a_0 + \cdots + \binom{m}{n+1} \Delta^n a_0 .$$

2. Furnish formulas for the sums $\sum_{\nu=0}^{m-1} \nu$, $\sum_{\nu=0}^{m-1} \nu^2$, $\sum_{\nu=0}^{m-1} \nu^3$.

23. FACTORIZATION

We saw already in Section 19 that the theorem on unique factorization holds for the polynomial domain $\mathsf{K}[x]$, where K is a commutative *field*. We shall proceed to prove the following more general *main theorem:*

If $\mathfrak{Z}$ is an integral domain with the identity, and if the unique factorization theorem holds in $\mathfrak{Z}$, then the same theorem holds for the polynomial domain $\mathfrak{Z}[x]$.

[2] For example, Kowalewski, *Interpolation und genäherte Quadratur* (Leipzig 1930); less detailed: R. Courant, *Differential and Integral Calculus* I. (London 1947), appendix to Chapter 6.

The proof is due to Gauss.

Let $f(x)=\sum_0^n a_i x^i$ be a polynomial in $\mathfrak{S}[x]$ distinct from zero. The greatest common divisor d of $a_0, \ldots, a_n$ in $\mathfrak{S}$ (cf. Section 19, Ex. 7) is called the *content* of $f(x)$. Factoring out d, we have

$$f(x)=d\cdot g(x),$$

where $g(x)$ has the content 1. $g(x)$ and d are uniquely determined, except for unit factors. Polynomials having content 1 are called *primitive polynomials* (with respect to $\mathfrak{S}$).

LEMMA 1. *The product of two primitive polynomials is itself primitive.*

PROOF: Let

$$f(x)=a_0+a_1x+\cdots$$

and

$$g(x)=b_0+b_1x+\cdots$$

be primitive polynomials. Let us suppose the coefficients of $f(x)\cdot g(x)$ have a common divisor d other than a unit. If p is a prime factor of d, then p must divide all coefficients of $f(x)g(x)$. Let a_r be the first coefficient of $f(x)$ not divisible by p (it must exist, otherwise $f(x)$ would not be a primitive polynomial); similarly, let b_s be the first coefficient of $g(x)$ not divisible by p.

The coefficient of x^{r+s} in $f(x)g(x)$ is of the form

$$a_rb_s+a_{r+1}b_{s-1}+a_{r+2}b_{s-2}+\cdots$$
$$+a_{r-1}b_{s+1}+a_{r-2}b_{s+2}+\cdots$$

The sum is supposed to be divisible by p. All terms except the first term are divisible by p. Hence a_rb_s must be divisible by p, i.e. either a_r or b_s has to be divisible by p, contrary to the assumption.

Let Σ be the quotient field of $\mathfrak{S}$ (Section 13). Then every polynomial in $\Sigma[x]$ is uniquely decomposable (Section 19). In order to pass from the decomposition in $\Sigma[x]$ to that in $\mathfrak{S}[x]$, we utilize the following fact: Every polynomial $\varphi(x)$ of $\Sigma[x]$ may be written in the form $\frac{F(x)}{b}$ ($F(x)$ in $\mathfrak{S}[x]$, b in $\mathfrak{S}$), where b is, say, the product of the denominators of the coefficients of $\varphi(x)$. Moreover, one may express $F(x)$ as the product of its "content by a primitive polynomial":

$$F(x)=a\cdot f(x),$$

$$\varphi(x)=\frac{a}{b}\cdot f(x). \tag{1}$$

Now we state the following:

LEMMA 2. *The primitive polynomial $f(x)$ appearing in (1) is uniquely determined by $\varphi(x)$ except for units in $\mathfrak{S}$. Conversely, $\varphi(x)$ is, according to (1), uniquely determined by $f(x)$, except for units in $\Sigma[x]$. Thus to any $\varphi(x)$ in $\Sigma[x]$ corresponds a primitive polynomial $f(x)$, and to the product of two polynomials $\varphi(x)$, $\psi(x)$ corresponds the product of the respective primitive polynomials (and vice versa). If $\varphi(x)$ is indecomposable in $\Sigma[x]$, so is $f(x)$ in $\mathfrak{S}[x]$ (and vice versa).*

PROOF: Let two different expressions for $f(x)$ be given:

$$\varphi(x) = \frac{a}{b} f(x) = \frac{c}{d} g(x).$$

Then

$$adf(x) = cbg(x). \tag{2}$$

follows.

The content of the left side is ad, that on the right side, cb; hence

$$ad = \varepsilon cb$$

where ε is a unit in $\mathfrak{S}$. Substituting it in (2) and dividing by cb, we get

$$\varepsilon f(x) = g(x).$$

Thus $f(x)$ and $g(x)$ differ from one another only by a unit in $\mathfrak{S}$.

For the product of two polynomials

$$\varphi(x) = \frac{a}{b} f(x),$$

$$\psi(x) = \frac{c}{d} g(x)$$

we obtain at once

$$\varphi(x) \cdot \psi(x) = \frac{ac}{bd} f(x) g(x).$$

By Lemma 1, $f(x)g(x)$ is again a primitive polynomial. Thus the product $f(x) \cdot g(x)$ corresponds to the product $\varphi(x) \cdot \psi(x)$.

If, finally, $\varphi(x)$ is indecomposable, so is $f(x)$; for a decomposition $f(x) = g(x)h(x)$ would immediately imply a decomposition

$$\varphi(x) = \frac{a}{b} f(x) = \frac{a}{b} g(x) \cdot h(x).$$

The converse can be proved in a similar fashion.

This completes the proof of Lemma 2.

By virtue of Lemma 2 the unique factorization of the polynomials $\varphi(x)$ may readily be applied to the respective primitive polynomials. Hence: *Primitive polynomials may uniquely (but for unit factors) be decomposed into prime factors which are themselves primitive polynomials.*

Let us now turn to the factorization of arbitrary polynomials in $\mathfrak{S}[x]$. Indecomposable polynomials are necessarily either indecomposable constants or indecomposable primitive polynomials; for all other polynomials can be factored into content times primitive polynomial. In order to factor a polynomial $f(x)$, it is first necessary to decompose it into content times primitive polynomial; then these two constituent parts have to be decomposed into prime factors separately. The first manipulation is, according to the main theorem, uniquely possible, and so is the second, according to what has just been proved. This *completes the proof* of the main theorem.

The following assertion is an additional result of the proof:

If a polynomial $F(x)$ in $\mathfrak{S}[x]$ is decomposable in $\Sigma[x]$, then it is already decomposable in $\mathfrak{S}[x]$.

For if we put $F(x)=d\cdot f(x)$, we obtain a primitive polynomial $f(x)$ corresponding to the polynomial $F(x)$, and according to Lemma 2 a factorization of $F(x)$ in $\Sigma[x]$ entails one of $f(x)$ in $\mathfrak{S}[x]$. Thus, if $f(x)$ is decomposable, so is $F(x)$.

For instance, a polynomial with rational integral coefficients which is decomposable in rational numbers is also decomposable in integral numbers. Therefore: *If an integral polynomial is indecomposable into factors with integral coefficients, it is also indecomposable into factors with rational coefficients.*

By complete induction we obtain another result from the main theorem:

If $\mathfrak{S}$ is an integral domain with an identity element, and if the unique factorization theorem is valid in $\mathfrak{S}$, then this theorem is likewise valid in the polynomial domain $\mathfrak{S}[x_1, \ldots, x_n]$.

From this theorem follows e.g. the unique factorization for polynomials with integer coefficients (in any number of variables), for polynomials with coefficients in a field, etc.

The concept of a "primitive polynomial," introduced in the Gaussian lemmas above, is particularly useful whenever we are dealing with polynomial domains in several variables. If K is a field, then a polynomial f of $\mathsf{K}[x_1, \ldots, x_n]$ is called *primitive with respect to* $x_1, \ldots, x_{n-1}$ if it is primitive with respect to the integral domain $\mathsf{K}[x_1, \ldots, x_{n-1}]$, i.e., if it does not have a non-constant factor that depends only on $x_1, \ldots, x_{n-1}$. For example, a polynomial is primitive with respect to $x_1, \ldots, x_{n-1}$, whenever it is "*regular with respect to* x_n," i.e., whenever the coefficient of the highest power of x_n is a constant distinct from zero (independent of $x_1, \ldots, x_{n-1}$).

EXERCISES. 1. The only units in $\mathfrak{S}[x]$ are those in $\mathfrak{S}$.

2. Prove that the factorization of a homogeneous polynomial yields only homogeneous factors.

3. Prove that the determinant

$$\Delta = \begin{vmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nn} \end{vmatrix}$$

is indecomposable in the polynomial domain $\mathfrak{S}[x_{11}, \ldots, x_{nn}]$. (Select one indeterminate, say x_{11}, and show that Δ is primitive with respect to the others.)

4. Establish a rule to decide whether a polynomial with integer coefficients has a factor of the first degree.

5. Prove the indecomposability of the polynomial

$$x^4 - x^2 + 1$$

in the polynomial domain of the indeterminate x over the ring of integers. Is the polynomial decomposable, when rational coefficients are allowed? Is it decomposable over the Gaussian ring?

24. IRREDUCIBILITY CRITERIA

Let $\mathfrak{S}$ be an integral domain with an identity element in which unique factorization holds. Let

$$f(x) = a_0 + a_1 x + \ldots + a_n x^n$$

be a polynomial in $\mathfrak{S}[x]$. The following theorem frequently supplies information as to the irreducibility of $f(x)$.

EISENSTEIN'S THEOREM: *If there exists a prime element p in $\mathfrak{S}$ such that*

$$\begin{aligned} a_n &\not\equiv 0 \,(p), \\ a_i &\equiv 0 \,(p) \quad \text{for all} \quad i < n, \\ a_0 &\not\equiv 0 \,(p^2) \end{aligned}$$

then $f(x)$ is irreducible in $\mathfrak{S}[x]$, except for constant factors; in other words, $f(x)$ is irreducible in $\Sigma[x]$, where Σ is the quotient field of $\mathfrak{S}$.

PROOF: Let us suppose $f(x)$ is decomposable, viz:

$$\begin{aligned} f(x) &= g(x) \cdot h(x), \\ g(x) &= \sum_0^r b_\nu x^\nu, \\ h(x) &= \sum_0^s c_\nu x^\nu, \\ r > 0, \quad s > 0, \quad r + s &= n, \end{aligned}$$

then we would have

$$a_0 = b_0 c_0 \quad \text{and} \quad a_0 \equiv 0 \,(p).$$

It follows that either $b_0 \equiv 0 \,(p)$ or $c_0 \equiv 0 \,(p)$. Let, for example, $b_0 \equiv 0 \,(p)$. Then $c_0 \not\equiv 0 \,(p)$, or else we would have $a_0 = b_0 c_0 \equiv 0 \,(p^2)$.

Not all the coefficients of $g(x)$ are divisible by p; for otherwise the product $f(x) = g(x) \cdot h(x)$ would be divisible by p, and all coefficients, in particular a_n, would be divisible by p, which contradicts the hypothesis. Thus let b_i be the first coefficient of $g(x)$ not divisible by p $(0 < i \leqq r < n)$.

Then

$$\begin{aligned} a_i &= b_i c_0 + b_{i-1} c_1 + \ldots + b_0 c_i, \\ a_i &\equiv 0 \,(p), \\ b_{i-1} &\equiv 0 \,(p), \\ &\ldots\ldots\ldots \\ b_0 &\equiv 0 \,(p), \end{aligned}$$

hence

$$\begin{aligned} b_i c_0 &\equiv 0 \,(p), \\ c_0 &\not\equiv 0 \,(p), \\ b_i &\equiv 0 \,(p), \end{aligned}$$

contrary to the hypothesis.

Hence $f(x)$ is irreducible, except for constant factors.

The criterion does not always lead to a decision; for there are many irreducible polynomials, such as $x^2 + 1$, to which it does not apply. Nevertheless, in favorable cases one obtains very general results from it.

EXAMPLE 1. $x^m - p$ (p prime) is irreducible over the ring of integers (and therefore also over the field of rational numbers). Hence $\sqrt[m]{p}$ ($m > 1$, p prime) is always irrational.

EXAMPLE 2. $f(x) = x^{p-1} + x^{p-2} + \cdots + 1$ is the left member of a "cyclotomic equation" if p is a prime number. We again ask for irreducibility over the ring of integers. The Eisenstein criterion cannot be applied directly; but we can reason as follows: If $f(x)$ were reducible, so would be $f(x+1)$. Now we have

$$f(x+1) = \frac{(x+1)^p - 1}{(x+1) - 1} = \frac{x^p + \binom{p}{1} x^{p-1} + \cdots + \binom{p}{p-1} x}{x}$$
$$= x^{p-1} + \binom{p}{1} x^{p-2} + \cdots + \binom{p}{p-1}.$$

All coefficients, except that of x^{p-1} are divisible by p; for in the formula for the binomial coefficients

$$\binom{p}{i} = \frac{p(p-1)\cdots(p-i+1)}{i!}$$

the numerator is divisible by p for $i < p$, but not the denominator. Furthermore, the constant term $\binom{p}{p-1} = p$ is not divisible by p^2. Hence $f(x+1)$ is irreducible, and so is $f(x)$.

EXAMPLE 3. For $f(x) = x^2 + 1$ the same transformation leads to a decision, since

$$f(x+1) = x^2 + 2x + 2 .$$

EXERCISES. 1. Prove the irrationality of $\sqrt[m]{p_1 p_2 \cdots p_r}$, where $p_1, \ldots, p_r$ are different prime numbers and $m > 1$.

2. Show that

$$x^2 + y^2 - 1$$

is irreducible in $\mathfrak{P}[x, y]$, where $\mathfrak{P}$ is any field in which $+1 \neq -1$.

3. Show that the polynomials

$$x^4 + 1, \quad x^6 + x^3 + 1$$

are irreducible in the integral polynomial domain.

Basically, the Eisenstein theorem rests on the fact that the equation

$$f(x) = g(x) \cdot h(x)$$

is transformed into a congruence modulo p^2, viz.

$$f(x) \equiv g(x) \cdot h(x),$$

which leads to an absurdity. In many other cases it is likewise possible to furnish irreducibility proofs by transforming the equations into congruences, modulo some quantity q of the domain $\mathfrak{S}$, and by investigating whether the polynomial $f(x)$ under consideration can be resolved modulo q. If, in particular, $\mathfrak{S}$ is the domain of the integers C, then there are only a finite number of polynomials of a given degree in the residue class domain modulo q; hence there are always but a finite number of possibilities of a resolution of $f(x)$ modulo q that have to be investigated. If it is found that $f(x)$ is irreducible modulo q, then $f(x)$ was also irreducible in $C[x]$, and even in the opposite case we might be able to draw conclusions from the decomposition modulo q. In the case where q is a prime number we may apply the unique factorization theorem of the polynomials modulo q (Section 19, Ex. 3).

EXAMPLE 1. $\mathfrak{S} = C$; $f(x) = x^5 - x^3 + 1$. If $f(x)$ modulo 2 is decomposable, then one of the factors has to be linear or quadratic. Now there are but two linear polynomials modulo 2:

$$x, \; x + 1,$$

and but one irreducible quadratic polynomial:

$$x^2 + x + 1.$$

On performing the division, we see that $x^5 - x^3 + 1$ is not divisible by any of these polynomials (modulo 2). This can be seen directly from

$$x^5 - x^3 + 1 = x^2(x^3 - 1) + 1 \equiv x^2(x+1)(x^2 + x + 1) + 1.$$

Hence $f(x)$ is irreducible.

EXAMPLE 2. $\mathfrak{S} = C$; $f(x) = x + 3x^3 + 3x^2 - 5$. $f(x)$ resolves modulo 2 into

$$f(x) \equiv (x+1)(x^3 + x + 1).$$

The last factor is irreducible modulo 2. Thus if $f(x)$ is resolvable, it is bound to resolve into a linear factor and a cubic factor. It is easy to show directly that a linear factor does not exist; this is done most conveniently by considering the fact that the only linear factors in question, viz, x, $x+1$, $x-1$ do not divide $f(x)$ modulo 3.

A far-reaching generalization of Eisenstein's irreducibility criterion is due to G. Dumas. Let p again be a prime element in $\mathfrak{S}$. To each term $ax^\lambda = cp^\mu x^\lambda$ of $f(x)$ distinct from zero with $(c, p) = 1$ there belongs a pair of exponents (λ, μ). We may think of these number pairs as the coordinates of as many lattice points in a (λ, μ)-plane as there are terms in $f(x)$.

We attach a *weight* $\alpha\lambda + \beta\mu$ to each term $cp^\mu x^\lambda$, where α, β are relatively prime integers, and where $\beta > 0$, i.e., we attach the weight α to the factors x, the positive weight β to the factors p, the weight 0 to the factors relatively prime to p, and we define the weight of a product as the sum of the weights of its factors.

Among all the weights of the terms of $f(x)$ there is a least value γ. Let us choose α and β so that this least value is assumed at least twice. The straight line $\alpha\lambda + \beta\mu = \gamma$ must therefore be so chosen that at least two of the lattice points under consideration are located on it and none of them below it. Then the quotient $\frac{-\alpha}{\beta}$ as a reduced fraction is the slope of the straight line.

Let (λ_1, μ_1) and (λ_2, μ_2) be two pairs of values of (λ, μ), for which $\alpha\lambda + \beta\mu$ assumes the least value γ. Choose λ_1 as small as possible and λ_2 as great as possible. From

$$\alpha\lambda_1 + \beta\mu_1 = \alpha\lambda_2 + \beta\mu_2 = \gamma$$

follows

(1) $$\alpha(\lambda_2 - \lambda_1) + \beta(\mu_2 - \mu_1) = 0;$$

hence $\alpha(\lambda_2 - \lambda_1)$ and therefore $(\lambda_2 - \lambda_1)$ are divisible by β.

$$\lambda_2 - \lambda_1 = m\beta, \quad \mu_2 - \mu_1 = -m\alpha, \quad m = (\lambda_2 - \lambda_1, \mu_2 - \mu_1).$$

We now state the following theorem: *If $f(x)$ is decomposable, the two factor polynomials have necessarily degrees of the form*

(2) $$m_1\beta + r_1 \quad \text{and} \quad m_2\beta + r_2$$

$$(m_1, m_2, r_1, r_2 \text{ are all } \geqq 0,\ m_1 + m_2 = m,\ r_1 + r_2 = n - m\beta).$$

PROOF. Let $f(x) = g_1(x) \cdot g_2(x)$, and let γ_1 be the least of the weights of the terms of $g_1(x)$, and γ_2 the least of the weights of the terms of $g_2(x)$. Among the terms of $g_1(x)$ of weight γ_1 let dx^δ be that with the least exponent δ, and let ex^ε be that with the largest exponent ε; accordingly, let rx^ϱ and sx^σ be defined for $g_2(x)$. On forming the product $g_1(x)\,g_2(x)$, we obtain some terms of weight $\gamma_1 + \gamma_2$; among them the term $drx^{\delta+\varrho}$ with the least, and the term $esx^{\varepsilon+\sigma}$ with the greatest exponent, while all other terms are of greater weights. The addition of terms of greater weights to a term such as $drx^{\delta+\varrho}$ or $esx^{\varepsilon+\sigma}$ of weight $\gamma_1 + \gamma_2$ does not alter this weight. Now, if the product $g_1 g_2$ is to be identical with $f(x)$, then, obviously, it is necessary that

$$\gamma_1 + \gamma_2 = \gamma, \quad \delta + \varrho = \lambda_1, \quad \varepsilon + \sigma = \lambda_2.$$

Hence it follows that

$$(\varepsilon - \delta) + (\sigma - \varrho) = \lambda_2 - \lambda_1 = m\beta.$$

Now, $\varepsilon - \delta$ and $\sigma - \varrho$ have to be divisible by β for the same reason $\lambda_2 - \lambda_1$ was, since δ and ε play the same role for $g_1(x)$ as λ_1 and λ_2 for $f(x)$. Hence

$$\varepsilon - \delta = m_1\beta, \quad \sigma - \varrho = m_2\beta, \quad m_1 + m_2 = m.$$

Finally, the degree of $g_1(x)$ is at least ε, hence $\geqq m_1\beta$, and that of $g_2(x)$ at least $m_2\beta$. Thus the above expressions for the degrees of $g_1(x)$ and $g_2(x)$ follow.

COROLLARIES. 1. *At least one of the degrees of* (2) *is* $\geqq \beta$.

2. *If the first and the last term of* $f(x)$ *are of last weight* γ, *then the degree of* g_1 *and* g_2 *are divisible by* β.

For in this case

$$m\beta = \lambda_2 - \lambda_1 = n - 0 = n, \quad r_1 + r_2 = n - m\beta = 0, \quad r_1 = r_2 = 0.$$

3. *If* $\beta = n$, *then* $f(x)$ *is irreducible. This follows from* 1.

In particular, if we take $\alpha = 1$, $\beta = \gamma = n$, we obtain Eisenstein's criterion.

EXAMPLES. 1. Let $f(x) = x^n + cp^m$, $(c, p) = 1$, $(m, n) = 1$. The linear form $m\lambda + n\mu$ has the value nm for both terms of $f(x)$. Thus we have to put $\alpha = m$, $\beta = n$. According to Corollary 3, $f(x)$ is indecomposable.

2. Let $n \geqq 2$ and $f(x) = x^n + px + bp^2$. The linear form $\lambda + (n-1)\mu$ has the value n for the first two terms, and $2n - 2 \geqq n$ for the last term. Thus we may set $\alpha = 1, \beta = n - 1$. If $f(x)$ is decomposable, then, by 1., one of the factors must be of degree $n - 1$, and the other one must therefore be linear.

25. FACTORIZATION IN A FINITE NUMBER OF STEPS

Until now we have only seen that there is a theoretical possibility to decompose into prime factors any polynomial in $\Sigma[x_1, \ldots, x_n]$ for a given field Σ, and in some instances we have provided the tools for actually furnishing a decomposition, or for showing the impossibility; yet, we still lack a general method for performing the factorization in a finite number of steps for any case that may present itself to us. We proceed to develop such a method at least for the case in which Σ is the field of rational numbers.

According to Section 23, we may assume the coefficients of any rational polynomial to be integers, and we may perform its factorization in the domain of integers. In the ring C of the integers itself a factorization into primes can evidently be performed by a finite trial and error method; furthermore, there are only a finite number of units ($+1$ and -1) in the ring C, and hence a finite number of possible factorizations. Similarly, in the polynomial domain $C[x_1, \ldots, x_n]$ there are only the units $+1, -1$. By the method of complete induction on the variable number n we shall now reduce everything to the following problem:

Let any factorization in $\mathfrak{S}$ *be performable in a finite number of steps: moreover, let there be only a finite number of units in* $\mathfrak{S}$. *We wish to find a method how to factor every polynomial in* $\mathfrak{S}[x]$ *into prime factors.*

The solution is due to Kronecker.

Let $f(x)$ be a polynomial of degree n in $\mathfrak{S}[x]$. If $f(x)$ can be factored, then one of the factors is of degree $\leqq \frac{n}{2}$; thus, if s is the greatest integer $\leqq \frac{n}{2}$, we have to investigate, whether $f(x)$ has a factor $g(x)$ of a degree $\leqq s$.

We form the function values $f(a_0), f(a_1), \ldots, f(a_s)$ for $s+1$ integral arguments $a_0, a_1, \ldots, a_s$. If $f(x)$ is to be divisible by $g(x)$, then $f(a_0)$ must be divisible by $g(a_0)$, and $f(a_1)$ by $g(a_1)$, etc. However, every $f(a_i)$ in $\mathfrak{S}$ possesses only a finite number of factors; therefore, for every $g(a_i)$ there are only a finite number of possibilities all of which may be found explicitly. For every possible combination of values $g(a_0), g(a_1), \ldots, g(a_s)$ there is, according to the theorems of Section 22, one, and only one, polynomial $g(x)$ which may be formed by Lagrange's or, more conveniently, Newton's interpolation formula. In this way a finite number of possible factors $g(x)$ are found. Employing the division algorithm, we may now find out whether each of these polynomials $g(x)$ is actually a factor of $f(x)$. If, apart from the units, none of the possible $g(x)$ is a factor of $f(x)$, then $f(x)$ is indecomposable; otherwise, a factorization has been found, and one may proceed to apply the same procedure to the two factors, etc. In this manner one finally arrives at the indecomposable factors.

In the integral case ($\mathfrak{S} = C$) the procedure may frequently be shortened considerably. By factoring the given polynomial modulo 2 and possibly modulo 3, we get an idea what degrees

the possible factor polynomials $g(x)$ might have, and what residue classes the coefficients modulo 2 and 3 might belong to. This limits the number of the possible $g(x)$ considerably. Moreover, when applying Newton's interpolation formula, one should note that the last coefficient λ_s must be a factor of the highest coefficient of $f(x)$, which limits the number of possibilities still further. Finally, it is an advantage to use more than $s+1$ points a_i (preferably $0, \pm 1, \pm 2$ etc.). For determining the possible $g(a_i)$ we use those $f(a_i)$ which contain the least number of prime factors; the other points may afterwards be used in order to limit the number of possibilities still further by examining each $g(x)$ and to see whether it assumes values which are factors of the respective $f(a_i)$ at all points a_i.

EXERCISES. 1. Factor

$$f(x) = x^5 + x^4 + x^2 + x + 2$$

in $C[x]$.

2. Eactor

$$f(x, y, z) = -x^3 - y^3 - z^3 + x^2(y+z) + y^2(x+z) + z^2(x+y) - 2xyz$$

in $C[x, y, z]$.

26. SYMMETRIC FUNCTIONS

Let $\mathfrak{o}$ be an arbitrary commutative ring with an identity element.

A polynomial in $\mathfrak{o}[x_1, \ldots, x_n]$ which is unchanged by any permutation of the indeterminates $x_1, \ldots, x_n$ is called a (rational integral) *symmetric function* of the variables $x_1, \ldots, x_n$. Examples: Sum, product, some of powers $s_\varrho = \sum_{\nu=1}^{n} x_\nu^\varrho$.

Introducing a new indeterminate z, we put

$$\begin{aligned} f(z) &= (z - x_1)(z - x_2)\ldots(z - x_n) \\ &= z^n - \sigma_1 z^{n-1} + \sigma_2 z^{n-2} - \cdots + (-1)^n \sigma_n . \end{aligned} \tag{1}$$

The coefficients of the powers of z in this polynomial are

$$\begin{aligned} \sigma_1 &= x_1 + x_2 + \cdots + x_n , \\ \sigma_2 &= x_1 x_2 + x_1 x_3 + \cdots + x_2 x_3 + \cdots + x_{n-1} x_n , \\ \sigma_3 &= x_1 x_2 x_3 + x_1 x_2 x_4 + \cdots + x_{n-2} x_{n-1} x_n , \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \sigma_n &= x_1 x_2 \ldots x_n . \end{aligned}$$

Obviously, they are all symmetric functions, since the left side of (1) remains unchanged by any permutations of the x_i. We call $\sigma_1, \ldots, \sigma_n$ the *elementary symmetric functions* of $x_1, \ldots, x_n$.

A polynomial $\varphi(\sigma_1, \ldots, \sigma_n)$ becomes a symmetric function of the $x_1, \ldots, x_n$ when the σ are written in terms of the x. Thus a term $c\sigma_1^{\mu_1} \ldots \sigma_n^{\mu_n}$ of $\varphi(\sigma_1, \ldots, \sigma_n)$ becomes a homogeneous polynomial in the x_i of degree $\mu_1 + 2\mu_2 + \cdots + n\mu_n$, since every σ is a homogeneous polynomial of the i-th degree. The sum $\mu_1 + 2\mu_2 + \cdots + n\mu_n$ will be called the *weight* of the term $c\sigma_1^{\mu_1} \ldots \sigma_n^{\mu_n}$. The weight of a polynomial $\varphi(\sigma_1, \ldots, \sigma_n)$ is defined as the largest weight occurring among its terms. Polynomials $\varphi(\sigma_1, \ldots, \sigma_n)$ of weight k, therefore, yield symmetric polynomials in the x_i of degree $\leqq k$.

The so-called *Fundamental Theorem on Symmetric Functions* asserts that the converse is also true:

A rational integral symmetric function of degree k in $\mathfrak{o}[x_1, \ldots, x_n]$ may be written as a polynomial $\varphi(\sigma_1, \ldots, \sigma_n)$ of weight k.

The most significant feature of this theorem is, of course, that any symmetric function may be expressed in terms of $\sigma_1, \ldots, \sigma_n$. The main purpose of introducing the ideas of degree and weight is to facilitate the proof by induction.

Before proceeding with the proof, we observe the following:

If two polynomials of $\mathfrak{o}[x_1, \ldots, x_n]$ are equal and if both are divisible by x_1, then the equality of the polynomials is preserved, when we cancel the factor x_1 from all the terms. This is true, regardless whether $\mathfrak{o}$ has zero divisors or not, since the equality of two polynomials simply implies complete equality of the coefficients.

If, in both members of the identical equation (1), we put $x_n = 0$, while the other x_i remain indeterminates, we get

$$(z - x_1) \cdots (z - x_{n-1})\, z = z^n - (\sigma_1)_0\, z^{n-1} + \cdots + (-1)^{n-1} (\sigma_{n-1})_0\, z,$$

where $(\sigma_i)_0$ is that expression which we obtain from σ_i for $x_n = 0$. If, in view of the above remark, we divide both sides by z, we obtain

$$(z - x_1) \cdots (z - x_{n-1}) = z^{n-1} - (\sigma_1)_0\, z^{n-2} + \cdots + (-1)^{n-1} (\sigma_{n-1})_0.$$

This equation means that the expressions $(\sigma_1)_0, \ldots, (\sigma_{n-1})_0$ are the elementary symmetric functions of the first $n - 1$ variables.

The proof of the *fundamental theorem* is given by induction on n. For $n = 1$ the theorem is true; for every polynomial $f(x_1)$ in one variable x_1 is symmetric, and $\sigma_1 = x_1$; hence $f(x_1) = f(\sigma_1)$. Assuming the theorem to hold for polynomials in $n - 1$ variables $(n > 1)$, we proceed to prove it for polynomials in n variables.

For polynomials of degree zero in n variables the theorem is trivial. Therefore, we may likewise assume that it has been proved for all polynomials of degrees $< k$ in n variables. It remains to be proved for polynomials of degree k in n variables.

Let a symmetric polynomial of the k-th degree $f(x_1, \ldots, x_n)$ be given. Taking $x_n = 0$, we have, by the induction assumptions,

$$f(x_1, \ldots, x_{n-1}, 0) = \varphi((\sigma_1)_0, \ldots, (\sigma_{n-1})_0),$$

where φ as a function of the elementary symmetric functions of $x_1, \ldots, x_{n-1}$ is of weight $\leqq k$. Therefore, the function $\varphi(\sigma_1, \ldots, \sigma_{n-1})$ has a weight $\leqq k$. Now form

$$f_1 = f(x_1, \ldots, x_{n-1}, x_n) - \varphi(\sigma_1, \ldots, \sigma_{n-1}).$$

The polynomial $f_1(x_1, \ldots, x_n)$ is obviously symmetric. The first term on the right-hand side is of degree k; the second is of weight $\leqq k$, and hence, as a polynomial

in the x, of degree $\leqq k$; consequently, f_1 is of degree $\leqq k$. Furthermore, for $x_n = 0$, f_1 vanishes; hence all terms contain the factor x_n. Since the function f_1 is symmetric, all the terms contain the factors $x_1, x_2, \ldots, x_{n-1}$.

Factoring out of all terms the product $x_1 x_2 \ldots x_n = \sigma_n$, we obtain

$$f_1 = \sigma_n g(x_1, \ldots, x_n),$$

where g is again a symmetric polynomial of degree $\leqq k-n<k$. Thus, by hypothesis, g may be expressed in terms of $\sigma_1, \ldots, \sigma_n$:

$$g = \psi(\sigma_1, \ldots, \sigma_n),$$

where ψ is a polynomial of weight $\leqq k-n$. Hence we may represent f as

$$f = f_1 + \varphi(\sigma_1, \ldots, \sigma_{n-1}) = \sigma_n \psi(\sigma_1, \ldots, \sigma_n) + \varphi(\sigma_1, \ldots, \sigma_{n-1}).$$

The right side represents a polynomial in σ of at most weight k. The weight cannot be less than k, since, otherwise, f would be of degree $< k$. Therefore the right-hand side is exactly of weight k. This completes the proof.

This proof provides a tool for expressing a given symmetric function in terms of the σ_i by actual computation. The method is, however, somewhat cumbersome, and we shall learn of a shorter one later on.

From the proof we further infer that homogeneous symmetric functions may be represented as "isobaric" expressions in σ_i, i.e. as expressions in which all terms have the same weight.

Let us now show that a symmetric function can be expressed as an integral rational function in $\sigma_1, \ldots, \sigma_n$ *in only one way*, or more precisely:

If $\varphi_1(y_1, \ldots, y_n)$ *and* $\varphi_2(y_1, \ldots, y_n)$ *are two polynomials in the indeterminates* $y_1, \ldots, y_n$, *and if*

$$\varphi_1(y_1, \ldots, y_n) \neq \varphi_2(y_1, \ldots, y_n),$$

then

$$\varphi_1(\sigma_1, \ldots, \sigma_n) \neq \varphi_2(\sigma_1, \ldots, \sigma_n).$$

If we form the difference $\varphi_1 - \varphi_2 = \varphi$, we see that it is sufficient to show that $\varphi(y_1, \ldots, y_n) \neq 0$ implies $\varphi(\sigma_1, \ldots, \sigma_n) \neq 0$.

The theorem holds for $n = 1$, since then $\sigma_1 = x_1$ is itself an indeterminate; hence $\varphi(y_1) \neq 0$ always implies $\varphi(\sigma_1) \neq 0$.

For an arbitrary $n > 1$ the theorem need be proved only under the assumption that it holds for any smaller number of indeterminates. Suppose the theorem does not hold for n; then there exists a polynomia $\varphi(y_1, \ldots, y_n) \neq 0$ of degree m as low as possible with respect to y_n so that $\varphi(\sigma_1, \ldots, \sigma_n) = 0$. If we arrange $\varphi(y_1, \ldots, y_n)$ in the order of the y_n, the two relations will be of the form

$$\varphi_m y_n^m + \varphi_{m-1} y_n^{m-1} + \cdots + \varphi_0 \neq 0,$$

$$\varphi_m(\sigma_1, \ldots, \sigma_{n-1})\, \sigma_n^m + \cdots + \varphi_0(\sigma_1, \ldots, \sigma_{n-1}) = 0. \tag{2}$$

It is necessary that $\varphi_0(y_1, \ldots, y_{n-1}) \neq 0$; otherwise y_n could be canceled from all the terms in the first relation, and σ_n in the second relation, and we would obtain:

$$\bar{\varphi}(y_1, \ldots, y_n) = \varphi_m y_n^{m-1} + \cdots + \varphi_1 \neq 0,$$

$$\bar{\varphi}(\sigma_1, \ldots, \sigma_n) = \varphi_m(\sigma_1, \ldots, \sigma_{n-1})\, \sigma_n^{m-1} + \cdots + \varphi_1(\sigma_1, \ldots, \sigma_{n-1}) = 0,$$

where the polynomial $\bar{\varphi}$ is of degree $< m$, contrary to hypothesis. Taking $x_n = 0$ in (2), we have

$$\varphi_0((\sigma_1)_0, \ldots, (\sigma_{n-1})_0) = 0,$$

although we had $\varphi_0(y_1, \ldots, y_{n-1}) \neq 0$, which contradicts the induction hypothesis. Thus we have proved the following:

A symmetric polynomial in $\mathfrak{o}[x_1, \ldots, x_n]$ *may be written in one, and only one, way as a polynomial in* $\sigma_1, \ldots, \sigma_n$; *the weight of this polynomial is equal to the degree of the given one.*

All integral rational relations between symmetric functions are preserved, if the x_i are not indeterminates, but are quantities in $\mathfrak{o}$, e.g., the roots of a polynomial $f(z)$ completely decomposable in $\mathfrak{o}[z]$. Thus it follows from what has been proved that a symmetric function of the roots of $f(z)$ may be expressed in terms of the coefficients of $f(z)$.

There are various methods for the computations necessary to express a given symmetric function as the elementary symmetric functions $\sigma_1, \ldots, \sigma_n$; only one of these methods will be presented here; others will be left as exercises. Arrange the given symmetric polynomial "*lexicographically*" (as in a dictonary), i.e., in such a fashion that a term $x_1^{\alpha_1} \ldots x_n^{\alpha_n}$ precedes another term $x_1^{\beta_1} \ldots x_n^{\beta_n}$ if the first non-vanishing difference $\alpha_i - \beta_i$ is positive. Together with one term $x_1^{\alpha_1} \ldots x_n^{\alpha_n}$ all terms occur whose exponents are a permutation of the α_i; we do not write all of them, but denote their sum by $\sum x_1^{\alpha_1} \ldots x_n^{\alpha_n}$, where it may be assumed that $\alpha_1 \geqq \alpha_2 \geqq \cdots \geqq \alpha_n$. Now for the initial term $a\, x_1^{\alpha_1} \ldots x_n^{\alpha_n}$ of the given polynomial we have to find a product of elementary symmetric functions which (when multiplied out and arranged lexicographically) has the same leading term $a\, x_1^{\alpha_1} \ldots x_n^{\alpha_n}$. This can be found easily; it is

$$a\, \sigma_1^{\alpha_1 - \alpha_2} \sigma_2^{\alpha_2 - \alpha_3} \ldots \sigma_n^{\alpha_n}.$$

This product is substracted from the given polynomial, and the terms are again arranged lexicographically. Next, the leading term is found again, etc.

EXERCISES. 1. Show that this method always yields the desired result, and derive from it a second proof for the fundamental theorem as well as for the uniqueness theorem.

2. For arbitrary n, express the "sums of powers" $\sum x_1$, $\sum x_1^2$, $\sum x_1^3$ by the elementary symmetric functions

3. Let $\sum x_1^\varrho = s_\varrho$. Prove the formulae

$$s_\varrho - s_{\varrho-1}\sigma_1 + s_{\varrho-2}\sigma_2 - \cdots + (-1)^{\varrho-1} s_1 \sigma_{\varrho-1} + (-1)^\varrho \varrho \sigma_\varrho = 0 \quad \text{for } \varrho \leqq n,$$

$$s_\varrho - s_{\varrho-1}\sigma_1 + \cdots + (-1)^n s_{\varrho-n}\sigma_n = 0 \quad \text{for } \varrho > n.$$

Employing these formulae, express the sums of the powers s_1, s_2, s_3, s_4, s_5 as the elementary symmetric functions.

4. Taking

$$s_\varrho = \sum a_{\lambda_1, \ldots, \lambda_n} \sigma_1^{\lambda_1} \sigma_2^{\lambda_2} \ldots \sigma_n^{\lambda_n}$$

by the fundamental theorem (summation for all λ_i with $\lambda_1 + 2\lambda_2 + 3\lambda_3 + \cdots = \varrho$), we obtain for the $a_{\lambda_1, \ldots, \lambda_n}$ (Ex. 3) the recursion formulae:

$$a_{\lambda_1, \ldots, \lambda_n} = a_{\lambda_1 - 1, \ldots, \lambda_n} - a_{\lambda_1, \lambda_2 - 1, \ldots, \lambda_n} + \cdots [+ (-1)^{\varrho - 1} \varrho],$$

where the term in square brackets occurs only when $\lambda_\varrho = 1$ (and then as the sole term), and when all other $\lambda_i = 0$, and where every a with a negative subscript is to bet set equal to zero. Prove that the solution of this recursive relation is:

$$a_{\lambda_1, \ldots, \lambda_n} = (-1)^{\lambda_2 + \lambda_4 + \lambda_6 + \ldots} \frac{\varrho \cdot (\lambda_1 + \lambda_2 + \cdots + \lambda_n - 1)!}{\lambda_1! \lambda_2! \cdots \lambda_n!}.$$

5. Let

$$(k_1, \ldots, k_h) = \sum x_1^{k_1} x_2^{k_2} \cdots x_h^{k_h}$$

with the summation performed on all *distinct* permuted terms which may be obtained if we take the order of the subscripts different from $1, 2, \ldots, h$. Prove that

$$\begin{aligned}(k_1, \ldots, k_h) \cdot (m) &= c_1 (k_1 + m, k_2, \ldots, k_h) \\ &+ c_2 (k_1, k_2 + m, \ldots, k_h) + \cdots \\ &+ c_h (k_1, k_2, \ldots, k_h + m) + c_0 (k_1, \ldots, k_h, m),\end{aligned}$$

where the coefficients $c_i (i = 1, \ldots, h)$ and c_0 indicate how many of the integers in the symbols to which they belong are equal to $k_i + m$ and to m, respectively.

6. Solve the formula found in Ex. 5 for $(k_1, \ldots, k_h, m)$, and derive from it a method of computation by means of which arbitrary symmetric functions may be expressed as the sums of powers (m) (provided that division by arbitrary integers distinct from zero is permissible in the underlying ring).

An important symmetric function is the square of the difference product:

$$D = \prod_{i<k} (x_i - x_k)^2.$$

The expression D, written as a polynomial in $a_1 = -\sigma_1$, $a_2 = \sigma_2, \ldots, a_n = (-1)^n \sigma_n$ is called the *discriminant* of the polynomial $f(z) = z^n + a_1 z^{n-1} + \cdots + a_n$. The vanishing of the discriminant for special $a_1, \ldots, a_n$ indicates that $f(z)$ has a multiple linear factor.

If we write out the polynomial $f(z)$ in a more general form with an arbitrary leading coefficient a_0, viz.

$$f(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_n,$$

then

$$\sigma_1 = -\frac{a_1}{a_0}, \qquad \sigma_2 = \frac{a_2}{a_0}, \ldots, \quad \sigma_n = (-1)^n \frac{a_n}{a_0}.$$

In this case we define the discriminant of $f(z)$ as the difference product multiplied by a_0^{2n-2}:

$$D = a_0^{2n-2} \prod_{i<k} (x_i - x_k)^2$$

In Section 28 we shall see that D is a polynomial in $a_0, a_1, \ldots, a_n$.

By employing the general method for expressing symmetric functions as polynomials in the coefficients, we find for the discriminants of $a_0 x^2 + a_1 x + a_2$:

$$D = a_1^2 - 4 a_0 a_2,$$

and of $a_0 x^3 + a_1 x^2 + a_2 x + a_3$:

$$D = a_1^2 a_2^2 - 4 a_0 a_2^3 - 4 a_1^3 a_3 - 27 a_0^2 a_3^2 + 18 a_0 a_1 a_2 a_3.$$

EXERCISE. 7. The discriminant remains invariant when every x_i is replaced by $x_i + h$. Derive from this fact the differential equation

$$\Omega D = n \frac{\partial D}{\partial a_1} + (n-1) a_1 \frac{\partial D}{\partial a_2} + \cdots + a_{n-1} \frac{\partial D}{\partial a_n} = 0.$$

27. THE RESULTANT OF TWO POLYNOMIALS

Let K be an arbitrary field, and let

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n,$$
$$g(x) = b_0 x^m + b_1 x^{m-1} + \cdots + b_m$$

be two polynomials in $\mathsf{K}[x]$. We wish to find a necessary and sufficient condition that the two polynomials have a non-constant common factor $\varphi(x)$.

We will not exclude the possibility that $a_0 = 0$ or $b_0 = 0$, i.e. that the degree of $f(x)$ is actually lower than n, or that the degree of $g(x)$ is lower than m. If the polynomial $f(x)$ is written in the above form, i.e. beginning with a (possibly vanishing) term $a_0 x^n$, then n is called the *formal degree* of the polynomial, and a_0 the *formal leading coefficient.* For the present we assume that at least one of the leading coefficients a_0, b_0 does not vanish.

Under this assumption we shall first show that $f(x)$ and $g(x)$ have a non-constant common divisor $\varphi(x)$ if; and only if, an equation of the form

$$h(x) f(x) = k(x) g(x) \tag{1}$$

exists, where $h(x)$ is at most of degree m—1, $k(x)$ at most of degree n—1, and where both polynomials h, k do not vanish identically.

If (1) is satisfied, and if we factor the two members of (1) into prime factors, we must obtain the same results both on the right and the left side. We may assume that $f(x)$, for example, is actually of degree n ($a_0 \neq 0$); for otherwise we need merely interchange rôles of $f(x)$ and $g(x)$. All prime factors of $f(x)$ must divide the right member of (1) just as often as $f(x)$. Yet they cannot divide $k(x)$ as often as they do $f(x)$; for $f(x)$ is at most of degree n—1. Hence at least one prime factor of $f(x)$ occurs also in $g(x)$. Q.E.D.

If, conversely, $\varphi(x)$ is a non-constant common factor of $f(x)$ and $g(x)$, it is merely necessary to put

$$f(x) = \varphi(x) k(x),$$
$$g(x) = \varphi(x) h(x),$$

and equation (1) will be satisfied.

In order to further investigate equation (1), we put

$$h(x) = c_0 x^{m-1} + c_1 x^{m-2} + \cdots + c_{m-1},$$
$$k(x) = d_0 x^{n-1} + d_1 x^{n-2} + \cdots + d_{n-1}.$$

The evaulation of equation (1) and a comparison of the coefficients of the powers $x^{n+m-1}, x^{n+m-2}, \ldots, x, 1$ on the left and on the right yields the following linear system of equations for the coefficients c_i and d_j:

$$(2)\quad \begin{cases} c_0 a_0 = d_0 b_0, \\ c_0 a_1 + c_1 a_0 = d_0 b_1 + d_1 b_0, \\ c_0 a_2 + c_1 a_1 + c_2 a_0 = d_0 b_2 + d_1 b_1 + d_2 b_0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ c_{m-2} a_n + c_{m-1} a_{n-1} = d_{n-2} b_m + d_{n-1} b_{m-1}, \\ c_{m-1} a_n = d_{n-1} b_m. \end{cases}$$

These are $n + m$ homogeneous linear equations for the $n + m$ quantities c_i, d_j. It is required that not all of these quantities vanish. A necessary condition for this is the vanishing of the determinant. In order to avoid minus signs in the determinant, we may regard the quantities c_i and $-d_j$ as unknowns, after transposing the right members of (2) to the left. Interchanging rows and columns (reflection at the principal diagonal), the determinant takes the form

$$(3)\quad R = \begin{vmatrix} a_0 & a_1 & \ldots & a_n & & & \\ & a_0 & a_1 & \ldots & a_n & & \\ & & \cdots & \cdots & \cdots & & \\ & & & a_0 & a_1 & \ldots & a_n \\ b_0 & b_1 & \ldots & b_m & & & \\ & b_0 & b_1 & \ldots & b_m & & \\ & & \cdots & \cdots & \cdots & & \\ & & & b_0 & b_1 & \ldots & b_m \end{vmatrix}$$

(In all blank spaces we have to substitute zeros.)

The determinant, as written out above, is called the *resultant* of the polynomials $f(x), g(x)$. We note that it is homogeneous of degree m in the a_i and homogeneous of degree n in the b_j; furthermore, the determinant contains the "principal term" $a_0^m b_m^n$ (principal diagonal), and, finally, it vanishes not only when the polynomials f, g have a common factor, but also when (contrary to the assumption made at the outset) $a_0 = b_0 = 0$.

Let us summarize:

The resultant of two polynomials $f(x)$, $g(x)$ is a rational integral form in the coefficients of the form (3). *If the resultant vanishes, the polynomials f and g have either a common non-constant factor, or the leading coefficient vanishes in both of them, and vice versa.*

The method of elimination used here was devised by Euler: the resultant of the form (3) is usually named after Sylvester.

The exception $a_0 = b_0 = 0$ may be discarded by taking two homogeneous

forms in two variables instead of two polynomials in one variable:

$$F(x) = a_0 x_1^n + a_1 x_1^{n-1} x_2 + \cdots + a_n x_2^n,$$
$$G(x) = b_0 x_1^m + b_1 x_1^{m-1} x_2 + \cdots + b_m x_2^m.$$

The original polynomials f, g and the numbers m, n determine the forms F, G uniquely, and vice versa. To a factorization of f, namely

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n$$
$$= (\rho_0 x^r + \cdots + \rho_r)(q_0 x^s + \cdots + q_s),$$

corresponds a factorization of F,

$$F(x) = a_0 x_1^n + \cdots + a_n x_2^n$$
$$= (\rho_0 x_1^r + \cdots + \rho_r x_2^r)(q_0 x_1^s + \cdots + q_s x_2^s),$$

and this is also true for g and G in a similar fashion. Hence to every common factor of f and g there corresponds a common factor of F and G. Conversely, if we put $x_1 = x$, $x_2 = 1$, every factorization of F and G yields at once a factorization of f and g, respectively, and every common factor of F and G yields a common factor of f and g. It may happen that the common factor of F and G is a pure power of x_2 and that, therefore, the common factor of f and g is a constant. But this case in which both F and G are divisible by x_2 is exactly the case $a_0 = b_0 = 0$; thus the two cases formulated above, in which the resultant vanishes, may be combined in one single statement: F and G possess a non-constant, homogeneous common factor.

We proceed to derive an important identity. Let the coefficients a_μ, b_ν of the polynomials $f(x)$, $g(x)$ now be indeterminates. We form

$$\begin{aligned}
x^{m-1} f(x) &= a_0 x^{n+m-1} + a_1 x^{n+m-2} + \cdots + a_n x^{m-1} \\
x^{m-2} f(x) &= \quad a_0 x^{n+m-2} + \cdots \quad + a_n x^{m-2} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
f(x) &= \quad a_0 x^n + \cdots \quad + a_n \\
x^{n-1} g(x) &= b_0 x^{n+m-2} + b_1 x^{n+m-2} + \cdots + b_m x^{n-1} \\
x^{n-2} g(x) &= \quad b_0 x^{n+m-2} + \cdots \quad + b_m x^{n-2} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
g(x) &= \quad b_0 x^m + \cdots \quad + b_m .
\end{aligned}$$

The determinant of this system of equations is exactly R. If we eliminate $x^{n+m-1}, \ldots, x$, on the right by multiplying by the subdeterminants of the last column, and by adding, we obtain an identity of the form [3]

$$Af + Bg = R, \tag{4}$$

where A and B are integral polynomials in the indeterminates a_μ, b_ν, x.

EXERCISES. 1. Give a determinant criterion for the fact that $f(x)$ and $g(x)$ have a factor in common at least of degree k.

2. For two polynomials of degree two we have

$$4R = (2a_0 b_2 - a_1 b_1 + 2a_2 b_0)^2 - (4a_0 a_2 - a_1^2)(4b_0 b_2 - b_1^2).$$

[3] For the forms F and G the corresponding relation is given by $AF + BG = x_2^{n+m-1} R$.

28. THE RESULTANT AS A SYMMETRIC FUNCTION OF THE ROOTS

We now assume that the two polynomials $f(x)$ and $g(x)$ *can be factored completely* into linear factors:

$$f(x) = a_0(x-x_1)(x-x_2)\dots(x-x_n)$$
$$g(x) = b_0(x-y_1)(x-y_2)\dots(x-y_m).$$

Then the coefficients a_μ of $f(x)$ are products of a_0 by the elementary symmetric functions of the roots $x_1, \dots, x_n$; similarly, the b_ν are products of b_0 by the symmetric functions of the y_k. The resultant R is homogeneous of degree m in the a_μ and homogeneous of degree n in the b_ν; hence R becomes equal to $a_0^m b_0^n$ times a symmetric function of the x_i and the y_k.

First, let the roots x_i and y_k be indeterminates. The polynomial R vanishes for $x_i = y_k$, since in this case the polynomials $f(x)$ and $g(x)$ have a linear factor in common. Hence R is divisible by $x_i - y_k$ (Section 19). Since the linear forms $x_i - y_k$ are relatively prime to each other, R must be divisible by the product

$$S = a_0^m b_0^n \prod_i \prod_k (x_i - y_k). \tag{1}$$

Now, this product may be transformed in two ways. Firstly, from

$$g(x) = b_0 \prod_k (x - y_k)$$

follows, upon substituting $x = \xi_i$ and forming a product, that

$$\prod_i g(x_i) = b_0^n \prod_i \prod_k (x_i - y_k);$$

hence

$$S = a_0^m \prod_i g(x_i). \tag{2}$$

Secondly, from

$$f(x) = a_0 \prod_i (x - x_i) = (-1)^n a_0 \prod_i (x_i - x)$$

follows in like manner that

$$S = (-1)^{nm} b_0^n \prod_k f(y_k). \tag{3}$$

It can be seen from (2) that S is integral and homogeneous of degree n in the b, and from (3) that S is integral and homogeneous of degree m in the a. But R has the same degrees as S and is divisible by S; hence R must coincide with S, except for a numerical factor. The comparison of those terms which contain the greatest power of b_m yields a term $+ a_0^m b_m^n$ both in R and in S; hence the value of the numerical factor is 1, and

$$R = S.$$

Thus we have found the three representations (1), (2), (3) for R. By the uniqueness theorem of Section 26, (2) holds identically in b_ν, and (3) identically in a_μ, i.e. (2) is valid even if $f(x)$ does not resolve into linear factors.

From the foregoing follows easily not only the *indecomposability of the resultant* as polynomial in the indeterminates $a_0, \ldots, b_m$ in an integral polynomial domain, but even its *absolute irreducibility*, i.e. the indecomposability in the polynomial domain of the same indeterminates with an arbitrary field as the domain of coefficients. For if R were decomposable into two factors A, B, then A and B could be written as symmetric functions of the roots. Since R is divisible by $x_1 - y_1$, A or B, say A, has to be divisible by $x_1 - y_1$ as well. But, being a symmetric function, A must be divisible by all other $x_i - y_k$ and therefore by their product

$$\prod_i \prod_k (x_i - y_k).$$

Since

$$R = a_0^m b_0^n \prod \prod (x_i - y_k),$$

there remains only one possibility for the other factor B, namely $B = a_0^p b_0^q$. But R as a polynomial in the a and b is divisible neither by a_0 nor by b_0; therefore, $B = 1$. This completes the proof of the irreducibility of R.

F. S. Macaulay gives a different proof in *Algebraic Theory of Modular Systems*. Section 3. (Cambridge 1916).

There is an interesting relationship between the resultant of two polynomials and the discriminant of a polynomial. Let us form the resultant $R(f, f')$ of the polynomial

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n = a_0 (x - x_1)(x - x_2) \cdots (x - x_n)$$

and its derivative $f'(x)$. By (2) we have

$$R(f, f') = a_0^{n-1} \prod f'(x_i). \tag{4}$$

According to the rules governing the differentiation of a product we have

$$f'(x) = \sum_i a_0 (x - x_1) \cdots (x - x_{i-1})(x - x_{i+1}) \cdots (x - x_n)$$

$$f'(x_i) = a_0 (x_i - x_1) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_n).$$

On substituting this in (4), we obtain

$$R(f, f') = a_0^{2n-1} \prod_{i \neq k} (x_i - x_k),$$

or, if D denotes the discriminant of $f(x)$:

$$R(f, f') = a_0 D. \tag{5}$$

Writing $R(f, f')$ as determinant according to Section 27, we may factor out the factor a_0 from the first column; therefore, D is a polynomial in $a_0, \ldots, a_n$. Once more, (5) holds identically in $a_0, \ldots, a_n$ whether or not $f(x)$ actually resolves into linear factors.

EXERCISES. 1. In the coefficients a and b together, the resultant of f and g is isobaric of weight mn (cf. Section 26).

2. If $y_1, \ldots, y_{n-1}$ are zeros of $f'(x)$, then

$$D = n^n a_0^{n-1} \prod_k f(y_k).$$

3. The discriminant D vanishes if, and only if, $f(x)$ and $f'(x)$ have a factor in common. If this is the case, then, upon factoring $f(x)$ into primes, we get either a multiple factor or a factor whose derivative vanishes identically.

29. PARTIAL FRACTION DECOMPOSITION

The following theorem on polynomials underlies the well-known partial fraction decomposition of rational functions: *If $g(x)$ and $h(x)$ are two relatively prime polynomials over a field K, if a is the degree of $g(x)$ and b that of $h(x)$, and if $f(x)$ is an arbitrary polynomial of degree less than $a + b$, then an identity*

(1) $$f(x) = r(x)\, g(x) + s(x)\, h(x)$$

exists, where $r(x)$ is of degree $< b$, and $s(x)$ of degree $< a$.

PROOF. By hypothesis, the greatest common divisor of $g(x)$ and $h(x)$ is equal to 1; therefore, the following identity holds:

$$1 = c(x)\, g(x) + d(x)\, h(x).$$

On multiplying both sides by f(x) we obtain

(2) $$f(x) = f(x)\, c(x)\, g(x) + f(x)\, d(x)\, h(x).$$

In order to reduce the degree of $f(x)c(x)$ to a value of $< b$, we divide this polynomial by $h(x)$:

(3) $$f(x)\, c(x) = q(x)\, h(x) + r(x),$$

where the degree of $r(x)$ is lower than that of $h(x)$ and thus lower than b. Substituting (3) in (2), it follows that

$$f(x) = r(x)\, g(x) + \{f(x)\, d(x) + q(x)\, g(x)\}\, h(x) = r(x)\, g(x) + s(x)\, h(x).$$

Here the left side and the first term on the right are of degree $< a + b$; hence the last term on the right is of degree $< a + b$, and therefore the degree of $s(x)$ is lower than a. This completes the proof of the above theorem.

If we divide both members of the identity (1) by $g(x)h(x)$, then the fraction $\frac{f(x)}{g(x)\,h(x)}$ resolves into two partial fractions

$$\frac{f(x)}{g(x)\,h(x)} = \frac{r(x)}{h(x)} + \frac{s(x)}{g(x)}.$$

By hypothesis, the degree of the numerator on the left is lower than that of the denominator, and the same is true of the two partial fractions on the right. If the denominator of one of these fractions can again be resolved into two relatively prime factors, then this fraction may again be resolved into two partial fractions.

We can proceed in this way, until the denominators have become powers of prime polynomials. Thus we may formulate the *theorem on partial fraction decomposition.*

A fraction $\frac{f(x)}{k(x)}$ *whose numerator is of degree lower than that of the denominator can be represented as a sum of partial fractions whose denominators are those powers of prime polynomials into which the denominator* $k(x)$ *resolves.*

The partial fractions $\frac{r(x)}{q(x)}$ thus obtained with the denominator $q(x) = p(x)^t$ may be split still further. If, for example, the prime polynomial $p(x)$ is of degree l and $q(x)$, therefore, of degree lt, then the numerator $r(x)$ whose degree is $< lt$ may first be divided by $p(x)^{t-1}$, leaving a remainder of degree $< l(t-1)$; then this remainder may be divided by $p(x)^{t-2}$, leaving a remainder of degree $< l(t-2)$, etc:

$$\begin{aligned} r(x) &= s_1(x)\, p(x)^{t-1} + r_1(x) \\ r_1(x) &= s_2(x)\, p(x)^{t-2} + r_2(x) \\ &\cdots\cdots \\ r_{t-2}(x) &= s_{t-1}(x)\, p(x) + r_{t-1}(x) \\ r_{t-1}(x) &= s_t(x). \end{aligned}$$

The quotients $s_1, \ldots, s_t$ are all of degree $< l$. From all these equations together follows that

$$r(x) = s_1(x)\, p(x)^{t-1} + s_2(x)\, p(x)^{t-2} + \cdots + s_{t-1}(x)\, p(x) + s_t(x)$$

$$\frac{r(x)}{p(x)^t} = \frac{s_1(x)}{p(x)} + \frac{s_2(x)}{p(x)^2} + \cdots + \frac{s_{t-1}(x)}{p(x)^{t-1}} + \frac{s_t(x)}{p(x)^t}. \tag{4}$$

Thus we have a second formulation of the theorem on partial fraction decomposition:

A fraction $\frac{f(x)}{k(x)}$ *whose numerator is of lower degree than that of the denominator and whose denominator has the factorization*

$$k(x) = p_1(x)^{t_1}\, p_2(x)^{t_2} \ldots p_h(x)^{t_h}$$

is a sum of partial fractions, the denominators being the powers $p_\nu(x)^{\mu_\nu}$ $(\mu_\nu = 1, 2, \ldots, t.;\ \nu = 1, 2, \ldots, h)$, *and the numerators being either zero, or having a lower degree than the prime polynomials* $p_\nu(x)$ *in the respective denominators.*

If, in particular, all the prime factors $p_\nu(x)$ are linear, the numerators of the partial fractions are constants. For this important special case we have a very simple method: By repeatedly splitting off a partial fraction with the highest possible exponent in the denominator, we can lower the degree of the denominator more and more. For if we write the denominator in the form $k(x) = (x-a)^t g(x)$, where $g(x)$ no longer contains the factor $x - a$, we have

$$\frac{f(x)}{k(x)} = \frac{f(x)}{(x-a)^t g(x)} = \frac{b}{(x-a)^t} + \frac{f(x) - b g(x)}{(x-a)^t g(x)}, \tag{5}$$

where the constant b can always be determined so that the numerator of the second fraction becomes zero for $x = a$ and is therefore divisible by $x - a$:

$$f(a) - bg(a) = 0$$

$$f(x) - b \cdot g(x) = (x - a) f_1(x).$$

In the second fraction in (5) the factor $x - a$ may now be canceled, and we may proceed to treat this fraction in a similar fashion, until the function is completely resolved into partial fractions.

CHAPTER V

THEORY OF FIELDS

The aim of this chapter is to give a general view of the structure of commutative fields, and of their simplest subfields and extention fields. Some of the subsequent investigations (Sections 30, 31, 33, 34) apply even to skew fields.

30. SUBFIELDS. PRIME FIELDS

Let Σ be a skew field.

If a subset Δ of Σ is itself a skew field, Δ is called a *subfield* of Σ. A necessary and sufficient condition for this is, first, that Δ be a subring (i.e., that together with a and b, it also contains $a - b$ and $a \cdot b$), and second, that it contain the identity and the inverse a^{-1} for every $a \neq 0$. Instead of this we may demand that Δ contain an element distinct from zero, and that, with a and b, it also contain $a - b$ and ab^{-1}.

It is obvious that *the intersection of any given number of subfields of Σ is itself a subfield of Σ.*

A *prime field* is a skew field that does not contain a proper subfield. We shall see afterwards that all prime fields are commutative.

Every skew field Σ contains one, and only one, prime field.

PROOF. The intersection of *all* subfields of Σ is a skew field which evidently has no proper subfields.

Let us suppose there exist two distinct prime subfields. Their intersection would be a subfield of both, and hence identical with both. Consequently the two prime fields would not be distinct from one another.

Types of prime fields. Let Π be the prime field contained in Σ. It contains the zero and the identity e, and therefore all integral multiples $n \cdot e = \pm \Sigma e$. Addition and multiplication of these elements are performed according to the followin rules:

$$ne + me = (n + m)\, e,$$
$$ne \cdot me = nm \cdot e^2 = nm \cdot e.$$

Thus the integral multiples ne form a commutative ring $\mathfrak{P}$. Furthermore, by $n \to ne$ a homomorphic mapping of a ring C of integers upon the ring $\mathfrak{P}$ is given. By the law of homomorphism (Section 16) $\mathfrak{P}$ is therefore isomorphic with a residue class ring $C/\mathfrak{p}$, where $\mathfrak{p}$ is the ideal of those integers n for which $ne = 0$. (In many fields $ne = 0$ is possible only for $n = 0$ so that $\mathfrak{p}$ is the null ideal. On the other hand, for example, in the field $C/(p)$ of residue classes modulo a prime number p, the equation $pe = 0$ holds.)

Since $\mathfrak{P}$ has no divisors of zero, $C/\mathfrak{p}$ cannot have any; hence $\mathfrak{p}$ must be a prime ideal. Moreover, $\mathfrak{p}$ cannot be the unit ideal; for in this case we should have $1 \quad e = 0$. Therefore there are two possibilities:

1. $\mathfrak{p} = (p)$, where p is a prime number. Then p is the least positive number such that $pe = 0$. We have in this case

$$\mathfrak{P} \cong C/(p).$$

$C/(p)$ is a field; therefore, the ring $\mathfrak{P}$ is a field and thus constitutes the prime field we were looking for. *Thus, in this case the prime field* Π *is isomorphic with the residue class ring modulo a prime number in the ring of integers. We operate with the elements* $n \cdot e$ *as we do with the residue classes of the numbers n mod p.*

2. $\mathfrak{p} = (0)$. The homomorphism $C \to \mathfrak{P}$ becomes a 1-isomorphism. Hence the multiples ne are all different: $ne = 0$ implies $n = 0$. In this case the ring $\mathfrak{P}$ is not yet a field; for the ring of integers is not a field. The prime field Π must contain not only the elements of $\mathfrak{P}$, but also their quotients. Now we know from Section 13 that the isomorphic integral domains $\mathfrak{P}$, C must also have isomorphic quotient fields so that in this case *the prime field* Π *is isomorphic with the field* Γ *of rational numbers.*

According to the foregoing, the structure of the prime field contained in Σ is completely defined by giving the number p or 0 which generates the ideal $\mathfrak{p}$. As was said before, $\mathfrak{p}$ consists of the numbers n such that $ne = 0$. The number p or 0 is called the *characteristic* of the skew field Σ or of the prime field Π.

All ordinary number and function fields which include the field of rational numbers are of characteristic zero.

The definition of the characteristic leads immediately to the following theorem:

Let $a \neq 0$ *be an element of* Σ, *and let k be the characteristic of* Σ. *Then* $na = ma$ *always implies* $n \equiv m(k)$, *and vice versa.*

PROOF. Multiplying the equation $na = ma$ by a^{-1}, we get $ne = me$, whence $n \equiv m(k)$ according to the definition of the characteristic. The conclusion is reversible.

We may prove in a similar fashion that $na = nb$ and $n \not\equiv 0(k)$ implies $a = b$.

Let us finally derive another important rule of operation:

In commutative fields of characteristic p we have

$$(a + b)^p = a^p + b^p,$$
$$(a - b)^p = a^p - b^p.$$

PROOF. In any commutative ring the binomial theorem (Section 11, Ex. 5)

$$(a+b)^p = a^p + \binom{p}{1} a^{p-1}b + \cdots + \binom{p}{p-1} a b^{p-1} + b^p$$

holds. But for $0 < i < p$: we have

$$\binom{p}{i} = \frac{p(p-1)\cdots(p-i+1)}{1\cdot 2\cdots i} \equiv 0(p),$$

since the numerator contains the factor p which cannot cancel. Hence, only the terms a^p and b^p remain:

$$(a+b)^p = a^p + b^p.$$

Substituting $a + b = a'$, we get

$$a'^p = (a'-b)^p + b^p,$$
$$(a'-b)^p = a'^p - b^p;$$

thus both assertions are proved.

EXERCISES. 1. Prove for characteristic p:

$$(a+b)^{p^f} = a^{p^f} + b^{p^f},$$
$$(a-b)^{p^f} = a^{p^f} - b^{p^f}$$

by the method of induction on f.

2. Similarly:

$$(a_1 + a_2 + \cdots + a_n)^p = a_1^p + a_2^p + \cdots + a_n^p.$$

3. Apply Ex. 2 to a sum $1+1+\cdots+1$ modulo p.

4. Prove for characteristic p:

$$(a-b)^{p-1} = \sum_{i=0}^{p-1} a^i b^{p-1-i}.$$

5. What are the characteristics of the residue class rings of the prime ideals $(1+i)$, (3), $(2+i)$ in the ring of the Gaussian integers (Section 17, Ex. 3)?

31. ADJUNCTION

If Δ is a subfield of a field Ω, then Ω is called an *extension field of* Δ. We want to get a general idea of all possible extensions of a given field Δ. This will give us at the same time a survey of all possible fields, since each field may be regarded as an extension of the prime field it contains. Our considerations must be limited to the commutative case; for non-commutative fields there does not exist a theory of this kind.

First of all, let Ω be a given extension field of Δ, and let $\mathfrak{S}$ be any set of elements in Ω. There are fields which include Δ and $\mathfrak{S}$. Ω, for example, is such a field. The intersection of all fields which include Δ and $\mathfrak{S}$ is itself a field including Δ and $\mathfrak{S}$ and will be denoted by $\Delta(\mathfrak{S})$. It is the smallest field that in-

cludes Δ and $\mathfrak{S}$. We say $\Delta(\mathfrak{S})$ arises from Δ by the *adjunction* (field adjunction) of the set $\mathfrak{S}$. We have

$$\Delta \subseteq \Delta(\mathfrak{S}) \subseteq \Omega,$$

and the two extreme cases are: $\Delta(\mathfrak{S}) = \Delta$, $\Delta(\mathfrak{S}) = \Omega$.

All elements of Δ, and all of $\mathfrak{S}$ belong to $\Delta(\mathfrak{S})$, and so do all those elements which arise from elements of Δ and $\mathfrak{S}$ by addition, subtraction, multiplication, and division. But these elements together form a field which, consequently, must be identical with $\Delta(\mathfrak{S})$. Hence, $\Delta(\mathfrak{S})$ *consists of all rational combinations of the elements of* $\mathfrak{S}$ *with those of* Δ. In the commutative case these combinations may be written simply as quotients of rational integral functions of the elements of $\mathfrak{S}$ with coefficients in Δ.

If $\mathfrak{S}$ is a finite set $\mathfrak{S} = \{u_1, \ldots, u_n\}$, we may write $\Delta(u_1, \ldots, u_n)$ instead of $\Delta(\mathfrak{S})$, and we speak of an adjunction of the elements $u_1, \ldots, u_n$ to Δ. The parentheses will, accordingly, always denote a field adjunction, whereas square brackets, e.g. $\Delta[x]$, will designate the ring adjunction (formation of sums and differences of products only).

In the rational expression of an element of $\Delta(\mathfrak{S})$ in terms of elements of Δ and of $\mathfrak{S}$ only a finite number of elements of $\mathfrak{S}$ can occur. Thus, every element of the field $\Delta(\mathfrak{S})$ already lies in a field $\Delta(\mathfrak{T})$, where $\mathfrak{T}$ is a finite subset of $\mathfrak{S}$. *Hence* $\Delta(\mathfrak{S})$ *is the union of all fields* $\Delta(\mathfrak{T})$, *where* $\mathfrak{T}$ *is each time a finite subset of* $\mathfrak{S}$. Thus the adjunction of an arbitrary set is reduced to adjunctions of finite sets and the formation of a union.

If $\mathfrak{S}$ is the union of $\mathfrak{S}_1$ and $\mathfrak{S}_2$, then, evidently,

$$\Delta(\mathfrak{S}) = \Delta(\mathfrak{S}_1)(\mathfrak{S}_2).$$

For $\Delta(\mathfrak{S}_1)(\mathfrak{S}_2)$ includes $\Delta(\mathfrak{S}_1)$ and $\mathfrak{S}_2$, and hence Δ, $\mathfrak{S}_1$ and $\mathfrak{S}_2$, or Δ and $\mathfrak{S}$, and therefore $\Delta(\mathfrak{S})$, and, conversely, $\Delta(\mathfrak{S})$ includes Δ, $\mathfrak{S}_1$ and $\mathfrak{S}_2$, hence $\Delta(\mathfrak{S}_1)$ and $\mathfrak{S}_2$, and therefore $\Delta(\mathfrak{S}_1)(\mathfrak{S}_2)$.

Thus it is seen that the adjunction of a finite set may be reduced to a finite number of successive adjunctions of a single element. Extensions by adjunction of a single element are called *simple field extensions*. Our first aim will be the study of such extensions.

32. SIMPLE FIELD EXTENSIONS

All fields to be considered in this section will be commutative. Again, let $\Delta \subseteq \Omega$, and let ϑ be an arbitrary element of Ω. Let us investigate the simple extension field $\Delta(\vartheta)$.

In the first place this field includes the ring $\mathfrak{S}$ of all polynomials $\sum a_k \vartheta^k$ $(a_k \in \Delta)$. We compare $\mathfrak{S}$ with the polynomial domain $\Delta[x]$ of an indeterminate x.

The mapping $f(x) \to f(\vartheta)$, or more precisely

$$\sum a_k x^k \to \sum a_k \vartheta^k$$

maps $\Delta[x]$ homomorphically upon $\mathfrak{S}$.[1] Thus, by the law of homomorphism, $\mathfrak{S}$ is isomorphic with a residue class ring

$$\mathfrak{S} \cong \Delta[x]/\mathfrak{p},$$

where $\mathfrak{p}$ is the ideal of those polynomials $f(x)$ which have ϑ as a root, i.e., for which $f(\vartheta) = 0$.

Since $\mathfrak{S}$ has no divisors of zero, $\Delta[x]/\mathfrak{p}$ cannot have any, either; hence the ideal $\mathfrak{p}$ must be prime. Furthermore, $\mathfrak{p}$ cannot be the unit ideal, since under the homomorphism the identity e is not associated with the zero but with e itself. Since every ideal in $\Delta[x]$ is a principal ideal, there remain but two possibilities:

1. $\mathfrak{p} = (\varphi(x))$, where $\varphi(x)$ is a polynomial irreducible in $\Delta[x]$.[2] $\varphi(x)$ is a polynomial of lowest degree with the property $\varphi(\vartheta) = 0$. It follows that

$$\mathfrak{S} \cong \Delta[x]/(\varphi(x)).$$

The residue class ring on the right is a field (Section 19); therefore, the ring $\mathfrak{S}$ is a field, and $\mathfrak{S}$ is the desired simple extension field $\Delta(\vartheta)$.

2. $\mathfrak{p} = (0)$. The homomorphism $\Delta[x] \sim \mathfrak{S}$ becomes an isomorphism. Except for the zero there is no polynomial $f(x)$ such that $f(\vartheta) = 0$, and we operate with the expressions $f(\vartheta)$ as if ϑ were an indeterminate x. In this case the ring $\mathfrak{S} \cong \Delta[x]$ is not a field yet; but the isomorphism of these rings implies the 1-isomorphism of their quotient fields: *The field $\Delta(\vartheta)$, quotient field of $\mathfrak{S}$, is isomorphic with the field of rational functions of an indeterminate x.*

In the first case, in which ϑ satisfies an algebraic equation $\varphi(\vartheta) = 0$ in Δ, ϑ is called *algebraic with respect to* Δ, and the field $\Delta(\vartheta)$ is called a *simple algebraic extension of* Δ. In the second case, in which $f(\vartheta) = 0$ implies $f(x) = 0$, ϑ is called *transcendental* with respect to Δ, and the field $\Delta(\vartheta)$ is called a *simple transcendental extension of* Δ. According to the above, we operate with a transcendental as we do with an indeterminate; we have $\Delta(\vartheta) \cong \Delta(x)$. In the algebraic case, however, we have according to the above:

$$\Delta(\vartheta) = \mathfrak{S} \cong \Delta[x]/(\varphi(x)),$$

where $\varphi(x)$ is the (irreducible) polynomial of lowest degree with ϑ as a root.

In the algebraic case the following facts follow from the last relation:

a) Every rational function of ϑ may be written as a polynomial $\sum a_k \vartheta^k$. (For $\mathfrak{S}$ was defined as the totality of these polynomials.)

---

[1] This is not true in the non-commutative case, since it has always been assumed that the variable x commutes with the coefficients a_k whereas ϑ need not commute with them. Only for the special case in which ϑ is interchangeable with all the elements of Δ do all considerations of this section apply.

[2] A less exact expression for "irreducible in $\Delta[x]$" is "irreducible in the field Δ" which is used occasionally. It might be better to say: "Irreducible over the field Δ."

b) We operate with these polynomials as we do with residue classes modulo $\varphi(x)$ in the polynomial domain $\Delta[x]$.

c) An equation

$$f(\vartheta)=0$$

may be transformed into a congruence

$$f(x)\equiv 0(\varphi(x))$$

and vice versa.

d) Since any polynomial $f(x)$ modulo $\varphi(x)$ may be reduced to a polynomial of degree $<n$, where n is the degree of $\varphi(x)$, all quantities of $\Delta(\vartheta)$ may be written in the form

$$\beta=\sum_{k=0}^{n-1} a_k \vartheta^k.$$

e) Since ϑ does not satisfy an equation of degree lower than the n-th, the representation

$$\beta=\sum_{k=0}^{n-1} a_k \vartheta^k$$

of the elements of $\Delta(\vartheta)$ is unique.

The irreducible equation $\varphi(x)=0$ having ϑ as a *root* is known as the *defining equation* of the field $\Delta(\vartheta)$. The degree of the polynomial $\varphi(x)$ is called the *degree* of the algebraic quantity ϑ with respect to Δ.

The degree is equal to 1 if ϑ is a solution of a *linear* equation in Δ, i.e., if ϑ itself belongs to the field Δ. In this case one may choose $\varphi(x)=x-\vartheta$. Thus the above theorem c) leads anew to the fact already proved in Section 21:

Every polynomial $f(x)$ having ϑ as a root is divisible by $x-\vartheta$.

EXERCISES. 1. For the case of a simple algebraic extension the irreducibility of the minimal polynomial $\varphi(x)$ as well as the statements a) to e) are to be proved directly, i.e., without using the law of homomorphism or the field properties of $\Delta[x]/(\varphi(x))$. [The order of the propositions is: Irreducibility, c), b), a), d), e). For a) use c).]

2. Also show that $\varphi(x)$ is, except for constant factors, the only polynomial irreducible in $\Delta[x]$ having ϑ as a root.

3. What is the degree of a generating element and the defining equation

a) of the field of complex numbers with respect to that of the real numbers,

b) of the field $\Gamma(\sqrt[5]{3})$ with respect to the field Γ of rational numbers,

c) of the field $\Gamma\left(e^{\frac{2\pi i}{5}}\right)$ with respect to the field Γ of rational numbers,

d) of the field $C[i]/(7)$ with respect to the prime field contained therein? ($C[i]$ is the ring of Gaussian integers.)

4. Let Γ be a commutative field, z an indeterminate, $\Sigma = \Gamma(z)$, $\Delta = \Gamma\left(\frac{z^3}{z+1}\right)$. Show that Σ is a simple algebraic extension of Δ. What equation, irreducible in Δ, is satisfied by the element z?

Two extensions Σ, Σ' of a field are said to be *equivalent* (with respect to Δ) if there exists a 1-isomorphism $\Sigma \cong \Sigma'$ which carries each element of Δ into itself, that is, which leaves each element fixed.

Any two simple transcendental extensions of a field Δ are equivalent.

For by means of $\frac{f(x)}{g(x)} \to \frac{f(\vartheta)}{g(\vartheta)}$ every simple transcendental extension $\Delta(\vartheta)$ is equivalent to the field of rational functions of the indeterminte x.

Two simple algebraic extensions $\Delta(\alpha)$, $\Delta(\beta)$ are equivalent as long as α and β are roots of the same polynomial $\varphi(x)$ irreducible in $\Delta[x]$; under this assumption there exists a 1-isomorphism which leaves the elements of Δ fixed and carries α into β.

PROOF: The elements of $\Delta(\alpha)$ are of the form $\sum_0^{n-1} a_k \alpha^k$, and those of $\Delta(\beta)$ are of the form $\sum_0^{n-1} a_k \beta^k$. In both cases we operate with these elements as we do with polynomials modulo $\varphi(x)$. The mapping

$$\sum a_k \alpha^k \to \sum a_k \beta^k$$

is seen to be an isomorphism of the kind desired.

A polynomial $\varphi(x)$ irreducible in Δ need not be irreducible in an extension field Ω. If it has a zero ϑ in Ω it splits off at least one linear factor $x - \vartheta$. It may be possible that it resolves into more linear or non-linear factors in Ω:

$$\varphi(x) = (x-\vartheta)(x-\vartheta_2)\cdots(x-\vartheta_j)\,\varphi_1(x)\cdots\varphi_k(x).$$

According to what was proved above, the fields $\Delta(\vartheta)$, $\Delta(\vartheta_2)$, ..., $\Delta(\vartheta_j)$ are all equivalent in this case, and under the isomorphisms

$$\Delta(\vartheta) \cong \Delta(\vartheta_2) \cong \cdots \cong \Delta(\vartheta_j)$$

ϑ is carried into $\vartheta_2, \ldots, \vartheta_j$.

Equivalent extensions [such as $\Delta(\vartheta)$, $\Delta(\vartheta_2)$, ..., $\Delta(\vartheta_j)$] which have a common extension field Ω are said to be *conjugate* with respect to Δ, and the elements $\vartheta, \vartheta_2 \ldots$, which are carried into each other under the respective 1-isomorphisms, are called *conjugate elements.*[3] From what was just proved follows: *All zeros in Ω of a polynomial $\varphi(x)$ irreducible in $\Delta[x]$ are conjugate among each other with respect to Δ.* Conversely, conjugate elements, if algebraic, are always roots of the same irreducible polynomial $\varphi(x)$; for if ϑ_1 is carried into ϑ_2 under a 1-isomorphism, $\varphi(\vartheta_1) = 0$ implies $\varphi(\vartheta_2) = 0$ by virtue of this very isomorphism.

[3] This term is mainly applied to algebraic elements ϑ. Transcendental elements of the same field are always conjugate among each other (see above).

THE EXISTENCE OF SIMPLE EXTENSIONS. Until now, Ω has always been a given extension field, and the structure of the simple extensions $\Delta(\vartheta)$ within Ω has been studied. We shall now pose the problem in a different manner: A field Δ is given, and an extension $\Delta(\vartheta)$ is to be found; moreover, it is required that ϑ be transcendental, or that it be a zero of a given polynomial irreducible in $\Delta[x]$.

If ϑ is to be transcendental, the solution is easy: We take for ϑ an indeterminate

$$\vartheta = x,$$

form the polynomial domain $\Delta[x]$ and its quotient field $\Delta(x)$, the field of rational functions of the indeterminate x. As we saw before, $\Delta(x)$ is the only simple transcendental extension, except for equivalent extensions; hence:

There exists one, and only one, simple transcendental extension $\Delta(\vartheta)$ of a given field Δ, except for equivalent extensions. Secondly, if ϑ is to be algebraic and a root of the polynomial $\varphi(x)$ irreducible in $\Delta[x]$, we may assume that φ is not linear, otherwise we could take $\Delta(\vartheta) = \Delta$.

According to the preceding paragraphs, the desired field $\Delta(\vartheta)$ must be isomorphic with the field of the residue classes

$$\Sigma' = \Delta[x]/(\varphi(x)).$$

Every polynomial f in $\Delta[x]$ defines a residue class $\bar{f}$ in Σ', and the mapping is homomorphic. In particular, to every constant a in Δ corresponds a residue class $\bar{a}$, and this mapping of Δ is not only homomorphic, but even 1-isomorphic, since zero is the only constant which is $\equiv 0 \bmod \varphi(x)$. According to what was said at the end of Section 12, we may, in the field Σ', replace the residue classes $\bar{a}$ by the corresponding elements a of Δ; in this way we obtain, instead of Σ', a field Σ which includes Δ, and which is $\cong \Sigma'$.

The polynomial x gives rise to a residue class which we shall call ϑ. Therefore, we may form the field $\Delta(\vartheta)$ in Σ. (It is easy to see that $\Sigma = \Delta(\vartheta)$, but we do not need this.) From

$$\varphi(x) = \sum_0^n a_k x^k \equiv 0 \,(\varphi(x))$$

follows by virtue of the isomorphism that

$$\sum_0^n \bar{a}_k \vartheta^k = 0 \text{ (in } \Sigma').$$

When the $\bar{a}_k$ are replaced by the a_k, it follows that

$$\varphi(\vartheta) = \sum_0^n a_k \vartheta^k = 0.$$

Hence ϑ is a root of $\varphi(x)$.

Thus we have proved:

For a given field Δ there exists one (and, except for equivalent extensions, only one) simple algebraic extension $\Delta(\vartheta)$ such that ϑ satisfies a given equation $\varphi(x) = 0$ irreducible in $\Delta[x]$.

The process of "symbolic adjunction" by means of the residue class ring and the symbol ϑ, as used in the proof, may be contrasted with the non-symbolic adjunction, which is possible if a comprehending field Ω is available at the outset which already contains a quantity ϑ with the required properties (cf. the beginning of this section). If, for example, Δ is the field of rationals, the non-symbolic adjunction of an alegbraic number, i.e., of a root of an algebraic equation may be attained by proceeding from the transcendentally constructed field Ω of complex numbers in which, by the "fundamental theorem of algebra," any equation with rational number coefficients is indeed solvable. In the above symbolic adjunction this transcendental detour is avoided by introducing the algebraic number as a symbol of a residue class directly, and by defining rules of operation for it. No inequality relations ($>$, $<$) or reality properties are introduced in this process. Nevertheless, both the symbolic and the transcendental method yield (algebraically speaking) the same field $\Delta(\vartheta)$; for, according to what was proved at the outset, all possible extensions $\Delta(\vartheta)$, with ϑ satisfying the same irreducible equation, are equivalent. The symbolic as well as the non-symbolic adjunction fall under the general concept of adjunction as outlined in Section 31. The only difference is that the comprehending field Ω or Σ necessary for the adjunction is in one case known beforehand, whereas in the other case it has to be constructed first.

More details regarding the connection between greater-and-smaller relations and algebraic relations will be found in Chapters IX and X.

EXERCISES. 5. The polynomial x^4+1 is irreducible in the field Γ of rationals (Section 24, Ex. 3). Adjoin a zero ϑ, and resolve the polynomial in the extended field $\Gamma(\vartheta)$ into prime factors.

6. Let Π be the prime field of characteristic p, let x be an indeterminate, and $\Delta=\Pi(x)$. Adjoin to Δ a root $\zeta=x^{\frac{1}{p}}$ of the irreducible polynomial z^p-x, and factor the polynomial z^p-x in the extended field $\Pi(\zeta)$.

7. From the prime field of characteristic 2 construct, by the adjunction of a root of an irreducible quadratic equation, a field with 4 elements.

33. LINEAR DEPENDENCE OVER A SKEW FIELD

Let $\mathfrak{M}$ be a module, i.e., an additive Abelian group containing the elements $u, v, \ldots$, and let Δ be a skew field with elements $\alpha, \beta, \ldots$.

Let a multiplication αv be defined (as in Section 14) such that

1. αv always belongs to $\mathfrak{M}$,
2. $\alpha(u+v)=\alpha u+\alpha v$,
3. $(\alpha+\beta)u=\alpha u+\beta u$,
4. $(\alpha\beta)u=\alpha(\beta u)$,
5. $1u=u$.

In the applications we are concerned with in this chapter $\mathfrak{G}$ is an extension field of Δ. However, we have deliberately chosen our hypotheses broad enough so that they may apply also to other additive groups, such as vector spaces, hypercomplex systems over Δ, as well as to arbitrary rings including Δ.

An element v of $\mathfrak{G}$ is called *linearly dependent on the elements* $u_1, \ldots, u_n$ (with respect to Δ) if

$$v = \alpha_1 u_1 + \cdots + \alpha_n u_n$$

or, what is the same thing, if there exists a linear relation

$$\beta_0 v + \beta_1 u_1 + \cdots + \beta_n u_n = 0$$

where $\beta_0 \neq 0$. In particular, v is called linearly dependent on the empty set if $v = 0$.

A number of theorems which, in the sequel, will be divided into "basic theorems" and "corollaries," is attached to the concept of linear dependence. The basic theorems will be deduced directly from the definition of the concept. The corollaries, on the other hand, will be derived from the basic theorems without using the definition again, i.e., without considering the meaning of the concept "linear dependence." Such a procedure (we may call it an axiomatization of the concept of linear dependence) is useful in view of a later chapter (Chapter VIII, Section 64), where the concept of "algebraic dependence" will be introduced for which the same basic theorems and, therefore, the same corollaries hold.

Three basic theorems will suffice. The first one is a matter of course.

FIRST BASIC THEOREM. *Every* u_i $(i = 1, \ldots, n)$ *is linearly dependent on* $u_1, \ldots, u_n$.

SECOND BASIC THEOREM. *If* v *is linearly dependent on* $u_1, \ldots, u_n$, *but not on* $u_1, \ldots, u_{n-1}$ *then* u_n *is linearly dependent on* $u_1, \ldots, u_{n-1}, v$.

PROOF. In the equation

$$\beta_0 v + \beta_1 u_1 + \cdots + \beta_n u_n = 0 \qquad (\beta_0 \neq 0)$$

it is necessary that $\beta_n \neq 0$, otherwise v would be dependent on $u_1, \ldots, u_{n-1}$.

THIRD BASIC THEOREM. *If* w *is linearly dependent on* $v_1, \ldots, v_s$ *and if every* v_j $(j = 1, \ldots, s)$ *is linearly dependent on* $u_1, \ldots, u_n$, *then* w *is linearly dependent on* $u_1, \ldots, u_n$.

PROOF. From $w = \sum_i \alpha_i v_i$ and $v_i = \sum_k \beta_{ik} u_k$ follows that

$$w = \sum_i \alpha_i \left(\sum_k \beta_{ik} u_k \right) = \sum_{i,k} \alpha_i \beta_{ik} u_k = \sum_k \left(\sum_i \alpha_i \beta_{ik} \right) u_k .$$

From the first and third basic theorems follows the

FIRST COROLLARY. *If* w *is linearly dependent on* $v_1, \ldots, v_s$, *then* w *is also linearly dependent on every system* $\{u_1, \ldots, u_n\}$ *which includes* $\{v_1, \ldots, v_s\}$.

A special case is that in which $v_1, \ldots, v_s$ coincides with $u_1, \ldots, u_n$, except for the order. Thus the concept of linear dependence is independent of the order of succession of $u_1, \ldots, u_n$.

DEFINITION. *The elements* $u_1, \ldots, u_n$ *are said to be linearly independent if none of them depends linearly on the others.*

This means (though we shall not make use of this meaning when deriving the corollaries) that we must have $\alpha_1 = 0, \ldots, \alpha_n = 0$ in every relation

$$\alpha_1 u_1 + \cdots + \alpha_n u_n = 0 .$$

The concept of linear independence does not depend on the order of $u_1, \ldots, u_n$. The empty set shall always be called linearly independent. A single element u is linearly independent if it does not depend on the empty set, *i.e.*, if $u \neq 0$.

SECOND COROLLARY. *If $u_1, \ldots, u_{n-1}$ are linearly independent, but $u_1, \ldots, u_{n-1}, u_n$ are not, then u_n is linearly dependent on $u_1, \ldots, u_{n-1}$.*

PROOF. Among the elements $u_1, \ldots, u_{n-1}, u_n$ one must be linearly dependent on the others. If this element is u_n, the proof is completed. If it is not u_n, but, say u_{n-1}, then u_{n-1} is linearly dependent on $u_1, \ldots, u_{n-2}, u_n$, but not on $u_1, \ldots, u_{n-2}$; hence (Second Basic Theorem) u_n is linearly dependent on $u_1, \ldots u_{n-2}, u_{n-1}$.

THIRD COROLLARY. *Any finite set of elements $u_1, \ldots, u_n$ contains a (possibly empty) linear independent set on which all u_i $(i = 1, \ldots, n)$ depend linearly.*

PROOF. Select a subset containing as many linearly independent elements as possible. Every u_i contained in the subset is, by the First Basic Theorem, linearly dependent on the subset, and so is every u_i not contained in the subset, by the Second Corollary.

DEFINITION. *Two finite sets $u_1, \ldots, u_n$ and $v_1, \ldots, v_s$ are said to be (linearly) equivalent if every v_k is linearly dependent $u_1, \ldots, u_n$, and every u_i on $v_1, \ldots, v_s$*

The equivalence definition is symmetric by definition; it is reflexive by the First Basic Theorem, and transitive by the Third Basic Theorem. If an element w is linearly dependent on one of the two equivalent sets, it is linearly dependent on the other one as well, according to the Third Basic Theorem. By the Third Corollary, every finite set is equivalent to the linearly independent subset.

FOURTH COROLLARY. (Replacement Theorem) *If $v_1, \ldots, v_s$ are linearly independent, and if every v_j is linearly dependent on $u_1, \ldots, u_n$, then there is in the set of the u_i a subset $\{u_{i_1}, \ldots, u_{i_s}\}$ of exactly s elements, which may be replaced by $\{v_1, \ldots, v_s\}$ so that the system obtained from $\{u_1, \ldots, u_n\}$ by this replacement is equivalent to the original set $\{u_1, \ldots, u_n\}$. This implies $s \leqq n$.*

PROOF. For $s = 0$ the proposition is trivial; for there are no v_j, and nothing is replaced. Now suppose the theorem holds for $\{v_1, \ldots, v_{s-1}\}$, and let $\{v_1, \ldots, v_{s-1}\}$ be replaceable by $\{u_{i_1}, \ldots, u_{i_{s-1}}\}$. This replacement gives rise to a set $\{v_1, \ldots, v_{s-1}, u_h, u_l, \ldots\}$ equivalent to $\{u_1, \ldots, u_n\}$. Now, v_s is linearly dependent on $\{u_1, \ldots, u_n\}$ and, therefore, on the equivalent set $\{v_1, \ldots, v_{s-1}, u_h, u_l, \ldots\}$. Thus, there is a minimal subset $\{v_1, \ldots, v_{s-1}, u_h, u_l, \ldots\}$, on which v_s still depends linearly. This minimal subset cannot consist of v_j only, since the v_j are linearly independent. Hence the minimal subset $\{v_j, \ldots, u_k\}$ con-

tains at least one u_k; which we shall call u_{i_s}. By the Second Fundamental Theorem $u_k = u_{i_s}$ is linearly dependent on the set which arises from $\{v_j, \ldots, u_k\}$ by replacing u_k by v_s and, therefore, on the including set which arises from $\{v_1, \ldots, v_{s-1}, u_k, u_l, \ldots\}$ by the replacement $u_k \to v_s$.

Let this system be $\{v_1, \ldots, v_{s-1}, v_s, u_l, \ldots\}$. It is equivalent to $\{v_1, \ldots, v_{s-1}, u_k, u_l, \ldots\}$ since u_k is linearly dependent on the first set, and v_s on the latter. In this manner we have carried the replacement one step further. The new set $\{v_1, \ldots, v_{s-1}, v_s, u_l, \ldots\}$ is equivalent to $\{v_1, \ldots, v_{s-1}, u_k, u_l, \ldots\}$, and, therefore, to the original system $\{u_1, \ldots, u_n\}$.

FIFTH COROLLARY. *Two equivalent linearly independent systems* $\{u_1, \ldots, u_r\}$ *and* $\{v_1, \ldots, v_s\}$ *consist of the same number of elements.*

PROOF. By the Fourth Corollary we have $s \leqq r$ and $r \leqq s$.

DEFINITION. *A set* $\mathfrak{M}$ *in* $\mathfrak{H}$ *is said to be of finite rank with respect to a skew field* Δ, *or briefly, finite over* Δ *if all elements of the set are linearly dependent on a finite number among them.* For example, a simple algebraic extension field $\Delta(\vartheta)$ is finite over Δ, since all elements of $\Delta(\vartheta)$ are, according to Section 32, linearly expressible in terms of $1, \vartheta, \vartheta^2, \ldots, \vartheta^{n-1}$. More generally speaking, every hypercomplex system over Δ is of finite rank.

According to the Third Corollary, we may select from among the finite number of elements on which all elements of $\mathfrak{M}$ are linearly dependent an equivalent linear independent subset $\{u_1, \ldots, u_r\}$. Such a linearly independent subset, on which all elements of $\mathfrak{M}$ depend linearly, is called a *basis*, or more precisely a *Δ-basis*[4] (also "minimal basis," or "linearly independent basis") of $\mathfrak{M}$.

If we express the elements of $\mathfrak{M}$ linearly in terms of the basis elements

$$u = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_r u_r, \tag{1}$$

the coefficients $\alpha_1, \ldots, \alpha_r$ will be uniquely determined; for

$$\alpha_1 u_1 + \cdots + \alpha_r u_r = \beta_1 u_1 + \cdots + \beta_r u_r$$

implies

$$\sum_i (\alpha_i - \beta_i) u_i = 0;$$

hence, because of the linear independence,

$$\alpha_i - \beta_i = 0.$$

Conversely, from the uniqueness of representation (1) follows the linear independence of the basis elements.

According to the Fifth Corollary, the number of basis elements is the same for each basis of $\mathfrak{M}$. This number is called the *linear rank of* $\mathfrak{M}$ with respect to Δ and is denoted by $(\mathfrak{M} : \Delta)$. In case of a finite extension field $\mathfrak{M}$ over Δ the linear

[4] The concept of Δ-basis should be distinguished from that of an ideal basis (Section 16). Either one is a special case of the more general concept of a module basis (basis of a module with respect to a domain of operators).

rank $(\mathfrak{M}:\Delta)$ is also called the *degree of the field* $\mathfrak{M}$ *over* Δ. In case of a vector space $\mathfrak{M}$, $(\mathfrak{M}:\Delta)$ is also said to be the *dimension of* $\mathfrak{M}$.

If r is the linear rank of $\mathfrak{M}$, then, according to the Fourth Corollary, the inequality $s \leqq r$ holds for every linearly independent system $v_1, \ldots, v_s$ in $\mathfrak{M}$. Therefore, the linear rank r may also be defined as the maximum number of linearly independent elements of the set. From this follows the

SIXTH COROLLARY. *A subset* $\mathfrak{N}$ *of a set* $\mathfrak{M}$ *finite with respect to* Δ *is itself finite, and the linear rank of* $\mathfrak{N}$ *is at most equal to that of* $\mathfrak{M}$.

EXERCISES. 1. If r is the rank of $\mathfrak{M}$, then any r linearly independent elements form a basis.

2. Every basis of a subset $\mathfrak{N}$ of $\mathfrak{M}$ can be extended to form a basis of $\mathfrak{M}$ if $\mathfrak{M}$ is of finite rank.

3. $\mathfrak{G} \geqq \Delta$ and $(\mathfrak{G}:\Delta) = 1$ imply $\mathfrak{G} = \Delta$.

We proceed to apply the concept of a linear rank to such extension fields of a field Δ as are finite over Δ. In commutative fields the linear rank $(\Sigma:\Delta)$ is generally called the *degree* of Σ over Δ. When the degree has the values 2, 3, 4, we speak of quadratic, cubic, biquadratic extension fields. A simple algebraic extension $\Sigma = \Delta(\vartheta)$ generated by an algebraic element ϑ of degree n is finite and of degree n, since the linearly independent powers $1, \vartheta, \vartheta^2, \ldots, \vartheta^{n-1}$ form a basis.

Let Σ be an intermediate field between Δ and Ω, i.e., let $\Delta \leqq \Sigma \leqq \Omega$. Then we have the following

THEOREM. *If* Ω *is finite over* Δ, *then* Σ *is finite over* Δ, *and* Ω *is finite over* Σ. *If, conversely,* Σ *is finite over* Δ *and* Ω *is finite over* Σ, *then* Ω *is finite over* Δ, *and we have the degree relation*

$$(\Omega:\Delta) = (\Omega:\Sigma)(\Sigma:\Delta). \tag{2}$$

PROOF. If Ω is finite over Δ, then Σ is finite over Δ by the Sixth Corollary. It is clear that Ω is finite over Σ, since Ω is finite even over Δ. Now let, conversely, $(\Sigma:\Delta)$ and $(\Omega:\Sigma)$ be finite, and let $\{u_1, \ldots, u_r\}$ be a basis of Σ with respect to Δ, and similarly $\{v_1, \ldots, v_s\}$ a basis of Ω with respect to Σ. Then every element of Ω can be represented in the form

$$\begin{aligned} w &= \sum_i \sigma_i v_i && (\sigma_i \in \Delta) \\ &= \sum_i \Big(\sum_k \delta_{ik} u_k\Big) v_i && (\delta_{ik} \in \Delta) \\ &= \sum_i \sum_k \delta_{ik} (u_k v_i). \end{aligned}$$

Thus, every element of Ω depends linearly on the rs quantities $u_k v_i$. These quantities are linearly independent of each other with respect to Δ; for since the v are linearly independent with respect to Σ, any linear dependence relation

$$\sum_i \sum_k \delta_{ik} u_k v_i = 0 \qquad (\delta_{ik} \in \Delta)$$

would imply

$$\sum_k \delta_{ik} u_k = 0;$$

hence, since the u are independent with respect to Δ,

$$\delta_{ik} = 0.$$

Hence rs is the degree of Ω with respect to Δ. Q.E.D.

Deductions from (2):

a) If $\Delta \subseteq \Sigma \subseteq \Omega$ and $(\Omega:\Delta) = (\Sigma:\Delta)$, then $\Omega = \Sigma$; for then it follows from (2) that $(\Omega:\Sigma) = 1$. Similarly:

b) If $\Delta \subseteq \Sigma \subseteq \Omega$ and $(\Omega:\Sigma) = (\Omega:\Delta)$, then $\Sigma = \Delta$.

c) If $\Delta \subseteq \Sigma \subseteq \Omega$, then the degree $(\Sigma:\Delta)$ is a factor of the degree $(\Omega:\Delta)$.

EXERCISES. 4. What is the degree of the field $\Gamma(i, \sqrt{2})$ with respect to the field Γ of rationals?

5. All elements of a finite commutative extension field Ω of a field Δ are algebraic with respect to Δ, and their degrees are factors of the degree $(\Omega:\Delta)$ of the field.

6. How many elements does a field of characteristic p contain if it is of degree n with respect to the prime field contained therein?

34. LINEAR EQUATIONS OVER A SKEW FIELD

An important application of the theory of linear dependence is the theory concerned with the solution of sets of linear equations.

Let $l_1, \ldots, l_m$ be linear forms in the indeterminates $x_1, \ldots, x_n$ with coefficients in a skew field Δ:

$$l_i = \sum \alpha_{ik} x_k. \tag{1}$$

Furthermore, let $\beta_1, \ldots, \beta_m$ be given elements in Δ. We wish to find (within Δ) all solutions $(\zeta_1, \ldots, \zeta_n)$ of the set of equations

$$l_i(\zeta) = \sum \alpha_{ik} \zeta_k = \beta_i \qquad (i = 1, 2, \ldots, m). \tag{2}$$

The number r of the linearly independent forms among the linear forms $l_1, \ldots, l_m$ is called the *rank* of the set of equations. Let us number the linear forms l_i so that $l_1, \ldots, l_r$ are linearly independent, and that the other l_i are linearly dependent on them:

$$l_i = \sum_1^r \varepsilon_{ik} l_k \qquad (i = r+1, \ldots, m). \tag{3}$$

The relations (3) remain valid if the x_j occurring in the l_i are replaced by ζ_j. For the set of equations (2) to be solvable it is necessary that the relations remain valid even when the l_i are replaced by β_i:

$$\beta_i = \sum_1^r \varepsilon_{ik}\beta_k \qquad (i = r+1, \ldots, m). \tag{4}$$

If these conditions are satisfied, all equations (2) are consequences of the first r among them. Therefore, we need merely consider these r independent equations.

The linear forms $l_1, \ldots, l_r$ are linearly dependent on the indeterminates $x_1, \ldots, x_n$. By the Replacement Theorem (Section 33, Fourth Corollary), we may replace r of these indeterminates, say (if numbered properly) $x_1, \ldots, x_r$, by $l_1, \ldots, l_r$ so that the new set $\{l_1, \ldots, l_r, x_{r+1}, \ldots, x_n\}$ is equivalent to the set $\{x_1, \ldots, x_n\}$. This means that all x are linearly dependent on $l_1, \ldots, l_r, x_{r+1}, \ldots, x_n$:

$$x_i = \sum_1^r \gamma_{ij} l_j + \sum_{r+1}^n \delta_{ik} x_k \qquad (i = 1, \ldots, r). \tag{5}$$

This relation, too, must remain valid when the x_i are replaced by ζ_i, and the l_j by β_j:

$$\zeta_i = \sum_1^r \gamma_{ij}\beta_j + \sum_{r+1}^n \delta_{ik}\zeta_k \qquad (i = 1, \ldots, r). \tag{6}$$

Thus the r unknowns $\zeta_1, \ldots, \zeta_r$ may be expressed linearly in terms of the other $\zeta_{r+1}, \ldots, \zeta_n$.

Until now we have set forth only the necessary conditions which have to be satisfied by all solutions of the given linear system of equations. We now assert that the necessary conditions are also sufficient.

If the conditions (4) are satisfied, the set of equations (2) is solvable, and the solutions are found by means of formula (6), where $\zeta_{r+1}, \ldots, \zeta_n$ may be chosen at will.

PROOF. Since the set $\{l_1, \ldots, l_r, x_{r+1}, \ldots, x_n\}$ is equivalent to the set $\{x_1, \ldots, x_n\}$, it is of the same rank n, and since it contains exactly n elements, these elements are linearly independent. Substituting (3) and (5) in (1), we obtain an identity in $l_1, \ldots, l_r, x_{r+1}, \ldots, x_n$, which is preserved when $l_1, \ldots, l_r$ are replaced by $\beta_1, \ldots, \beta_r$, and $x_{r+1}, \ldots, x_n$ by arbitrary elements $\zeta_{r+1}, \ldots, \zeta_n$. This means that the β_i determined from (4) and the ζ_i determined from (6) satisfy equations (2).

For actually determining the rank r, for finding the linearly independent l_i, and for establishing the solution formula (6) we use (in practice also) the method of *successive elimination*. First, we solve one of the relations (2) for an x_j, then substitute this x_j in the other relations (2), thus replacing a basis element x_j by an l_i. We continue in this manner until no more x will remain in the expressions for the remaining l_i, and thus these l_i depend (let us say) on $l_1, \ldots, l_r$ alone. Then we can find out whether the linear dependences (3) thus found hold for the β_i as well, or, what is more convenient, we replace the l_i at once by the known

β_i from the very beginning. The formulae (5) or (after replacing the l by the β and the x by the ζ) the formulae (6) follow automatically from the successive substitutions.

From the possibility of this rational computation follows: *If the coefficients of a set of linear equations belong to a subfield of* Δ , *then the coefficients of formulae* (6) *lie in the same subfield. If the set is solvable in* Δ, *it is already solvable in the subfield.*

In the special case of a set of *homogeneous* equations (all $\beta_i = 0$) the necessary conditions (3) are automatically satisfied. Thus, the set of homogeneous equations is always solvable, and the solutions are given by (6), where $\beta_j = 0$. In this case, (5) may also be interpreted as follows: All vector solutions $(\zeta_1, \ldots, \zeta_n)$ are linear combinations of n—r special vector solutions

$$\begin{aligned}
&(\delta_{1r+1}, \delta_{2r+1}, \ldots, \delta_{rr+1}, 1, 0, \ldots, 0)\\
&(\delta_{1r+2}, \delta_{2r+2}, \ldots, \delta_{rr+2}, 0, 1, \ldots, 0)\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&(\delta_{1n}, \quad \delta_{2n}, \ldots, \quad \delta_{rn}, \quad 0, 0, \ldots, 1)
\end{aligned}$$

and are obtained from these by multiplying them in order, *on the right*, by the arbitrary elements $\zeta_{r+1}, \ldots, \zeta_n$ and adding. In the special case $r = n$ there is only the "trivial solution" $(0, \ldots, 0)$.

In case of a commutative field Δ, the theory of determinants furnishes explicit solution formulae and algebraic criteria for the solvability and linear dependence of linear equations. For this we refer the reader to the text-books.

35. ALGEBRAIC FIELD EXTENSIONS

An extension field Σ of Δ is called *algebraic over* Δ if every element of Σ is algebraic over Δ .

THEOREM. *Every finite extension* Σ *of* Δ *is algebraic, and may be obtained from* Δ *by the adjunction of a finite number of algebraic elements.*

PROOF. If n is the degree of the finite extension Σ , and $\alpha \in \Sigma$, then the powers $1, \alpha, \alpha^2, \ldots, \alpha^n$ of an element α contain at most n linearly independent ones. Therefore, a relation $\sum_0^n c_k \alpha^k = 0$ must exist, i.e., α is algebraic; consequently the field Σ is algebraic. As generator of the extension Σ (i.e., as adjoined set) we may choose a field basis of Σ.

By virtue of this theorem we may say "finite algebraic extension" instead of "finite extension".

CONVERSE. *Every extension of a field* Δ *obtained by the adjunction of a finite number of algebraic quantities to* Δ *is finite (and therefore algebraic).*

PROOF. The adjunction of an algebraic quantity ϑ of degree n yields a

finite extension with the basis $1, \vartheta, \ldots, \vartheta^{n-1}$. By the last theorem of Section 33, successive finite extensions always yield a finite extension.

COROLLARY. *The sum, difference, product, and quotient of algebraic quantities are themselves algebraic quantities.*

Theorem. *If α is algebraic with respect to Σ, and if Σ is algebraic with respect to Δ, then α is algebraic with respect to Δ.*

PROOF. In the algebraic equation for α with coefficients in Σ only a finite number of elements $\beta, \gamma, \ldots$ of Σ can occur as coefficients. The field $\Sigma' = \Delta(\beta, \gamma, \ldots)$ is finite with respect to Δ, and the field $\Sigma'(\alpha)$ is finite with respect to Σ'; hence $\Sigma'(\alpha)$ is also finite with respect to Δ and therefore α is algebraic with respect to Δ.

DECOMPOSITION FIELDS. Among the finite algebraic extensions, the "*decomposition fields*" of a polynomial $f(x)$ which are obtained by the "*adjunction of all roots of an equation* $f(x)=0$" are of special significance. A decomposition field is a field $\Delta(\alpha_1, \ldots, \alpha_n)$, in which the polynomial $f(x)$ in $\Delta[x]$ completely resolves into linear factors [5]

$$f(x) = (x-\alpha_1)\cdots(x-\alpha_n),$$

and which are obtained by the adjunction of the roots α_i of these linear factors to Δ. The following theorems apply to such fields:

For every polynomial $f(x)$ in $\Delta[x]$ there exists a decomposition field.

PROOF. Let $f(x)$ be resolved in $\Delta[x]$ into prime factors as follows:

$$f(x) = \varphi_1(x)\varphi_2(x)\cdots\varphi_r(x).$$

We first adjoin a root α_1 of the irreducible polynomial $\varphi_1(x)$ and thus obtain a field $\Delta(\alpha_1)$ in which $\varphi_1(x)$ and therefore $f(x)$ splits off a linear factor $x-\alpha_1$.

Let us suppose we have already constructed a field $\Delta_k = \Delta(\alpha_1, \ldots, \alpha_k)$ $(k<n)$ in which the polynomial $f(x)$ splits off the (equal or different) factors $x-\alpha_1, \ldots, x-\alpha_k$. Let $f(x)$ be resolved in the field Δ_k as follows:

$$f(x) = (x-\alpha_1)\cdots(x-\alpha_k)\cdot\psi_{k+1}(x)\cdots\psi_l(x).$$

We now adjoin to Δ_k a root α_{k+1} of $\psi_{k+1}(x)$. In the field

$$\Delta_k(\alpha_{k+1}) = \Delta(\alpha_1, \ldots, \alpha_{k+1})$$

thus extended $f(x)$ splits off the factors $x-\alpha_1, \ldots, x-\alpha_{k+1}$. It could happen but it would not matter that, after the adjunction, $f(x)$ might spilt off even more than these $k+1$ linear factors. Continuing in this manner step by step, we eventually find the desired field $\Delta_n = \Delta(\alpha_1, \ldots, \alpha_n)$.

We proceed to show that the decomposition field of a given polynomial $f(x)$ is uniquely determined, except for equivalent extensions. For this purpose we must familiarize ourselves with the concept of the *continuation of an isomorphism*.

[5] Here and in the sequel we shall assume the highest coefficient of $f(x)$ to be 1, which obviously is not essential.

Let $\Delta \subseteq \Sigma$ and $\bar{\Delta} \subseteq \bar{\Sigma}$, and let a 1-isomorphism $\Delta \cong \bar{\Delta}$ be given. A 1-isomorphism $\Sigma \cong \bar{\Sigma}$ is called a *continuation* of the given 1-isomorphism $\Delta \cong \bar{\Delta}$, if every quantity a of Δ which, under the original isomorphism $\Delta \cong \bar{\Delta}$ is mapped upon $\bar{a}$, has the same image $\bar{a}$ in $\bar{\Delta}$ under the new isomorphism $\Sigma \cong \bar{\Sigma}$.

All theorems on continuations of isomorphisms in algebraic extensions are based on the following theorem:

If, under a 1*-isomorphism* $\Delta \cong \bar{\Delta}$, *an irreducible polynomial* $\varphi(x)$ *in* $\Delta[x]$ *is carried into a polynomial* $\bar{\varphi}(x)$ (*which of course is likewise irreducible*) *in* $\bar{\Delta}[x]$, *and if* α *is a root of* $\varphi(x)$ *in an extension field of* Δ, *and* $\bar{\alpha}$ *a root of* $\bar{\varphi}(x)$ *in an extension field of* $\bar{\Delta}$, *then the given* 1*-isomorphism* $\Delta \cong \bar{\Delta}$ *may be continued to a* 1*-isomorphism* $\Delta(\alpha) \cong \bar{\Delta}(\bar{\alpha})$, *which carries* α *into* $\bar{\alpha}$.

PROOF: The elements of $\Delta(\alpha)$ are of the form $\sum c_k \alpha^k (c_k \in \Delta)$, and we operate with them just as with polynomials modulo $\varphi(x)$. Similarly, the elements of $\bar{\Delta}(\bar{\alpha})$ are of the form $\sum \bar{c}_k \bar{\alpha}^k (\bar{c}_k \in \bar{\Delta})$, and we operate with them just as with polynomials modulo $\bar{\varphi}(x)$, i.e., exactly so, except for the horizontal bars. Therefore, the mapping

$$\sum c_k \alpha^k \to \sum \bar{c}_k \bar{\alpha}^k$$

(where the c_k correspond to the c_k under the isomorphism $\Delta \cong \bar{\Delta}$) is an isomorphism having the required properties.

If, in particular, $\Delta = \bar{\Delta}$, and if the given isomorphism maps every element of Δ upon itself, we obtain the previous theorem again, according to which all extensions $\Delta(\alpha), \Delta(\bar{\alpha}), \ldots$, each of which arises by the adjunction of the root of the same irreducible equation, are equivalent.

A similar theorem holds for the adjunction of all roots of a polynomial instead of only one:

If under a 1*-isomorphism* $\Delta \cong \bar{\Delta}$ *an arbitrary polynomial* $f(x)$ *in* $\Delta[x]$ *is carried into a polynomial* $\bar{f}(x)$ *in* $\bar{\Delta}[x]$, *then the* 1*-isomorphism may be continued to a* 1 *isomorphism of an arbitrary decomposition field* $\Delta(\alpha_1, \ldots, \alpha_n)$ *of* $f(x)$ *with an arbitrary decomposition field* $\bar{\Delta}(\bar{\alpha}_1, \ldots, \bar{\alpha}_n)$ *of* $\bar{f}(x)$, *with* $\alpha_1, \ldots, \alpha_n$ *being carried into* $\bar{\alpha}_1, \ldots, \bar{\alpha}_n$ *in a certain order.*

PROOF. Let us suppose we have already continued (possibly after changing the order of the roots) the 1-isomorphism $\Delta \cong \bar{\Delta}$ to a 1-isomorphism

$$\Delta(\alpha_1, \ldots, \alpha_k) \cong \bar{\Delta}(\bar{\alpha}_1, \ldots, \bar{\alpha}_k),$$

mapping every α_i upon $\bar{\alpha}_i$. (For $k = 0$ this is trivial.) In $\Delta(\alpha_1, \ldots, \alpha_k)$ let $f(x)$ be decomposed thus:

$$f(x) = (x - \alpha_1) \cdots (x - \alpha_k) \cdot \varphi_{k+1}(x) \cdots \varphi_k(x).$$

By applying the 1-isomorphism, we find that $\bar{f}(x)$ is decomposed in $\bar{\Delta}(\bar{\alpha}_1, \ldots, \bar{\alpha}_k)$ as follows:

$$\bar{f}(x) = (x - \bar{\alpha}_1) \cdots (x - \bar{\alpha}_k) \cdot \bar{\varphi}_{k+1}(x) \cdots \bar{\varphi}_k(x).$$

Furthermore, in $\Delta(\alpha_1, \ldots, \alpha_n)$ and $\bar{\Delta}(\bar{\alpha}_1, \ldots, \bar{\alpha}_n)$ respectively, the factors φ_r and $\bar{\varphi}_r$ resolve into $(x - \alpha_{k+1}) \cdots (x - \alpha_n)$ and $(x - \bar{\alpha}_{k+1}) \cdots (x - \bar{\alpha}_n)$. Let the $\alpha_{k+1}, \ldots \alpha_n$

and $\bar{\alpha}_{k+1}, \ldots, \bar{\alpha}_n$ be rearranged in such a way that α_{k+1} becomes the root of $\varphi_{k+1}(x)$, and $\bar{\alpha}_{k+1}$ that of $\bar{\varphi}_{k+1}(x)$. By the previous theorem the 1-isomorphism

$$\Delta(\alpha_1, \ldots, \alpha_k) \cong \bar{\Delta}(\bar{\alpha}_1, \ldots, \bar{\alpha}_k)$$

may be continued to an isomorphism

$$\Delta(\alpha_1, \ldots, \alpha_{k+1}) \cong \bar{\Delta}(\bar{\alpha}_1, \ldots, \bar{\alpha}_{k+1})$$

which maps α_{k+1} upon $\bar{\alpha}_{k+1}$.

Starting from $k=0$ and proceeding step by step in this way, we finally arrive at the desired 1-isomorphism

$$\Delta(\alpha_1, \ldots, \alpha_n) \cong \bar{\Delta}(\bar{\alpha}_1, \ldots, \bar{\alpha}_n),$$

which maps every α_i upon $\bar{\alpha}_i$.

If, in particular, $\Delta = \bar{\Delta}$, and if the given 1-isomorphism $\Delta \cong \bar{\Delta}$ leaves every element of Δ fixed, then $\bar{f} = f$, and the extended 1-isomorphism

$$\Delta(\alpha_1, \ldots, \alpha_n) \cong \bar{\Delta}(\bar{\alpha}_1, \ldots, \bar{\alpha}_n)$$

again leaves all elements of Δ fixed, i.e., the two decomposition fields of $f(x)$ are *equivalent. Therefore, the decomposition field of a polynomial $f(x)$ is uniquely determined, except for equivalent extensions.*

From this it follows that all algebraic properties of the roots are independent of the method of construction of the decomposition field. For instance, regardless whether a polynomial is decomposed in the field of complex numbers, or by means of a symbolic adjunction, we always get "essentially," i.e., except for equivalence, the same result.

In particular, every root or zero of $f(x)$ has a definite *multiplicity* in which it occurs in the decomposition

$$f(x) = (x-\alpha_1) \cdots (x-\alpha_n)$$

Multiple roots exist when, and only when, $f(x)$ and $f'(x)$ have a non-constant common divisor over the decomposition field (Section 21). But the greatest common divisor of $f(x)$ and $f'(x)$ over any decomposition field is the same as the greatest common divisor in $\Delta[x]$ (Section 18, Ex. 1). Thus, by forming the greatest common divisor of $f(x)$ and $f'(x)$ in $\Delta[x]$, we can see beforehand whether $f(x)$ will have multiple roots in its decomposition field.

Two decomposition fields of one and the same polynomial which are contained in a common comprehending field Ω are not only equivalent but even *equal.* For if two decompositions

$$f(x) = (x-\alpha_1) \cdots (x-\alpha_n),$$
$$f(x) = (x-\bar{\alpha}_1) \cdots (x-\bar{\alpha}_n)$$

take place in Ω , then, by the unique factorization theorem in $\Omega[x]$, the factors coincide, except for their order.

NORMAL EXTENSION FIELDS. A field Σ is called *normal* over Δ if, first, it is algebraic with respect to Δ , and secondly, every polynomial $g(x)$ irreduc-

ible in $\Delta[x]$, which has *one* root α in Σ, is completely decomposed into linear factors in $\Sigma[x]$.[6]

Our previously constructed decomposition fields are normal according to the following theorem:

A field which arises from Δ by the adjunction of all zeros of one, or several, or even infinitely many polynomials in $\Delta[x]$ is normal.

In the first place we may reduce the case of an infinite number of polynomials to that of finitely many; for each element α of the field depends solely on the roots of a finite number of our polynomials, and for the decomposition of the irreducible polynomial which has α as zero we may confine ourselves entirely to the field generated by this finite number of roots..

Then we may reduce the case of a number of polynomials to that of a single one by multiplying together all of them, and by adjoining the roots of the product.

Thus, let $\Sigma=\Delta(\alpha_1, \ldots, \alpha_n)$, where the α_ν are the roots of a polynomial $f(x)$, and let the irreducible polynomial $g(x)$ in $\Delta[x]$ have a root β in Σ. If $g(x)$ is not completely decomposed in Σ, we can extend Σ to a field $\Sigma(\beta')$ by the adjunction of another root β' of $g(x)$; then, since β and β' are conjugate, we have

$$\Delta(\beta) \cong \Delta(\beta').$$

In this isomorphism the elements of Δ and, in particular, the coefficients of the polynomial $f(x)$ remain fixed. If we adjoin all roots of $f(x)$ on the left and on the right, the isomorphism may be continued:

$$\Delta(\beta, \alpha_1, \ldots, \alpha_n) \cong \Delta(\beta', \alpha_1, \ldots, \alpha_n),$$

where the α_i are again mapped upon the α_j, perhaps in a different order. Now β is a rational function of $\alpha_1, \ldots, \alpha_n$ with coefficients in Δ,

$$\beta = r(\alpha_1, \ldots, \alpha_n),$$

and this rational relation is preserved in every isomorphism. Therefore, β' is also a rational function of $\alpha_1, \ldots, \alpha_n$, and thus also belongs to the field Σ, contrary to our assumption.

CONVERSE. *A normal field Σ over Δ arises by the adjunction of all roots of a set of polynomials and, if the field is finite, it arises even by the adjunction of all zeros of one single polynomial.*

PROOF. Let the field Σ be generated by the adjunction of a set $\mathfrak{M}$ of algebraic elements. (In the general case we may, for example, take $\mathfrak{M}=\Sigma$; in

[6] Another version of the definition is the following: *An algebraic extension Σ is normal if Σ contains, together with any elements α, all elements* (of any comprehending field) conjugate to α. For the elements conjugate to α in any comprehending field are nothing else but the roots of the same irreducible polynomial $g(x)$ having α as a root, and the comprehending field may be chosen so that $g(x)$ is completely decomposed in it. This definition, however, has been avoided, since it refers to the totality of all comprehending fields. Even apart from the doubtfulness of the concept of such a totality from a set-theoretical viewpoint, this definition appears less elegant, since we are actually concerned with a property only of Σ and Δ.

the finite case $\mathfrak{M}$ is finite.) Every element of $\mathfrak{M}$ satisfies an algebraic equation $f(x)=0$ with coefficients in Δ which is decomposed completely in Σ. The adjunction of all roots of all these polynomials $f(x)$ (or, if their number is finite, the adjunction of all roots of their product) yields at least as much as the adjunction of $\mathfrak{M}$ alone, i.e., it yields the entire field Σ. Q.E.D.

An irreducible equation $f(x)=0$ is called *normal* if the field obtained by the adjunction of *one* root is already normal, i.e., if $f(x)$ is completely decomposed in it.

By a *Galois resolvent* of an equation $f(x)=0$ is meant an irreducible equation $g(x)=0$ such that the adjunction of one root of this equation yields already the entire decomposition field of the polynomial $f(x)$. The existence of such resolvents will be proved later (Section 40).

EXERCISES. 1. If $\Delta \subseteq \Sigma \subseteq \Omega$, and if Ω is normal over Δ, then Ω is normal over Σ.

2. Construct the decomposition field of x^3-2 with respect to the rational field Γ. Show that, if α is any root, $\Gamma(\alpha)$ is not normal.

3. If $f(x)$ is irreducible in the field K, then in a normal extension field all prime factors of $f(x)$ are of the same degree, and are conjugate with respect to K.

4. Every field quadratic with respect to Δ is normal with respect to Δ.

36. ROOTS OF UNITY

In the preceding sections we presented the fundamentals of the theory of fields in general. Before continuing to develop the general theory, we shall apply it to some very special equations and special fields.

Let Π be a prime field, and let h be a natural number which is not congruent to zero modulo the characteristic of Π. (If the characteristic is zero, h may be any natural number.) By an *h-th root of unity* we shall mean a root of the polynomial

$$f(x)=x^h-1$$

in any commutative extension field.

The h-th roots of unity in a field form an Abelian group under multiplication.

For if $\alpha^h=1$ and $\beta^h=1$, then $\left(\frac{\alpha}{\beta}\right)^h=1$, from which the group property follows. It is obvious that the group is an Abelian group.

The order of a group element α is a divisor of h, since we must have $\alpha^h=1$.

The decomposition field Σ of $f(x)$ is called *the field of the h-th roots of unity over the prime field* Π. The polynomial $f(x)$ is decomposed into linear factors which are all different from each other; for the derivative

$$f'(x)=hx^{h-1}$$

never vanishes, since h is not divisible by the characteristic, except if $x=0$, and

therefore has no root in common with $f(x)$. *Thus there are exactly h h-th roots of unity in* Σ.

We now resolve h into prime factors:

$$h = \prod_{i=1}^{m} q_i^{\nu_i}.$$

In the group of the *h-th* roots of unity there are at most $\frac{h}{q_i}$ elements a, for which $a^{\frac{h}{q_i}} = 1$; for the polynomial $x^{\frac{h}{q_i}} - 1$ has at most $\frac{h}{q_i}$ roots. Therefore, there is an a_i in the group with

$$a_i^{h:q_i} \neq 1.$$

The group element

$$b_i = a_i^{h:q_i^{\nu_i}}$$

is of order $q_i^{\nu_i}$; for its $q_i^{\nu_i}$-th power is 1, and its order is therefore a divisor of $q_i^{\nu_i}$; but its $q_i^{\nu_i - 1}$-st power is different from 1; hence its order is not a proper divisor of $q_i^{\nu_i}$. The product

$$\zeta = \prod_{1}^{m} b_i$$

is a product of elements of relatively prime order $q_1^{\nu_1}, \ldots, q_m^{\nu_m}$; hence its order is exactly

$$\prod_{1}^{m} q_i^{\nu_i} = h$$

(Section 7, Ex. 1). If the order of a root of unity is exactly h, it will be called *a primitive h-th root of unity.*

The powers $1, \zeta, \zeta^2, \ldots, \zeta^{h-1}$ of a primitive root of unity are all different; but since the group does not contain more than h elements, all group elements are powers of ζ. Thus:

The group of the h-th roots of unity is cyclic and is generated by every primitive root of unity ζ.

It is easy to determine the number of primitive h-th roots of unity. For the present we denote it by $\varphi(h)$. *$\varphi(h)$ is the number of the elements of order h in a cyclic group of order h.*[7] First, if h is a power of a prime number, $h = q^\nu$, all q^ν powers of ζ, excepting the $q^{\nu-1}$ powers of ζ^q are elements of order h. Hence

$$\varphi(q^\nu) = q^\nu - q^{\nu-1} = q^{\nu-1}(q-1) = q^\nu\left(1 - \frac{1}{q}\right). \tag{1}$$

Secondly, if h is decomposed into two relatively prime factors $h = rs$, every element of order h is uniquely representable as the product of an element of order r by an element of order s (Section 18, Ex. 3) and, conversely, every such product is an element of order h. The elements of the r-th order belong to the cyclic group of order r generated by ζ^s; their number is $\varphi(r)$. Similarly, the number of the

[7] According to Section 18, Ex. 4, $\varphi(h)$ is also the number of the natural numbers $\leq h$ relatively prime to h. $\varphi(h)$ is called Euler's φ-function.

elements of order s is $\varphi(s)$; thus, for the number of the products we have

$$\varphi(h) = \varphi(r)\varphi(s).$$

If, as before,

$$h = \prod_1^m q_i^{\nu_i}$$

is the decomposition of h into relatively prime powers of prime numbers, the above formula yields by repeated application

$$\varphi(h) = \varphi(q_1^{\nu_1} q_2^{\nu_2} \cdots q_m^{\nu_m}) = \varphi(q_1^{\nu_1})\varphi(q_2^{\nu_2}) \cdots \varphi(q_m^{\nu_m});$$

hence by (1):

$$\begin{aligned}\varphi(h) &= q_1^{\nu_1 - 1}(q_1 - 1)\, q_2^{\nu_2 - 1}(q_2 - 1) \cdots q_m^{\nu_m - 1}(q_m - 1) \\ &= h\left(1 - \frac{1}{q_1}\right)\left(1 - \frac{1}{q_2}\right) \cdots \left(1 - \frac{1}{q_m}\right).\end{aligned}$$

Thus we have:

The number of the primitive h-th roots of unity is

$$\varphi(h) = h \prod_1^m \left(1 - \frac{1}{q_i}\right).$$

We put $n = \varphi(h)$. Let the primitive h-th roots of unity be $\zeta_1, \ldots, \zeta_n$. They are the roots of the polynomial

$$(x - \zeta_1)(x - \zeta_2) \cdots (x - \zeta_n) = \Phi_h(x).$$

We have

$$x^h - 1 = \prod_{d|h} \Phi_d(x), \tag{2}$$

where d runs over the positive divisors of h;[8] for every h-th root of unity is a primitive d-th root of unity for one, and only one, positive divisor d of h, and therefore every linear factor of $x^h - 1$ occurs in one, and only one, of the polynomials $\Phi_d(x)$.

Formula (2) determines $\Phi_h(x)$ uniquely. First, (2) implies

$$\Phi_1(x) = x - 1,$$

and, if Φ_d is known for all positive $d < h$, Φ_h can be determined from (2) by division.

Since, by the division algorithm, these divisions can be performed in the domain of polynomials in x with integer coefficients, we have:

Every $\Phi_h(x)$ is a polynomial with integer coefficients, independent of the characteristic of the field Π (if only h is not divisible by it).

For a reason to be mentioned later, these polynomials $\Phi_h(x)$ are known as

[8] $a|b$ (in words: a divides b) means: a is a divisor of b.

cyclotomic (*circle-dividing*) *polynomials*. Explicit rational formulae may be given for them by means of the "Möbius Function" $\mu(n)$, which is defined as follows:

$$\mu(n) = \begin{cases} 0 & \text{if } p_i^2 | n \text{ for any } p_i, \\ (-1)^\lambda & \text{if } n = p_1 p_2 \dots p_\lambda \text{ (i.e., if } n \text{ is "square-free"),} \\ 1 & \text{if } n = 1 ; \end{cases}$$

($p_1, \dots, p_\lambda$ are the various prime factors of the number n). The Möbius function has the following important property:

$$\sum_{d|h} \mu(d) = \begin{cases} 1 & \text{for } h = 1, \\ 0 & \text{for } h > 1. \end{cases}$$

This property may be proved, e.g., by putting $h = q_1^{\nu_1} \dots q_m^{\nu_m}$, developing the product $\prod_1^m (1 - z_i)$, and then making all z_i equal to 1. The terms

$$(-1)^\lambda z_{i_1} z_{i_2} \dots z_{i_\lambda}$$

of this product correspond exactly to the non-quadratic divisors $d = q_{i_1} q_{i_2} \dots q_{i_\lambda}$ of h, and we have

$$(-1)^\lambda = \mu(d).$$

Thus, taking all $z_i = 1$, we get for $m > 0$ (i.e. $h > 1$),

$$0 = \prod_1^m (1 - 1) = \sum_{d|h} \mu(d),$$

while for $h = 1$ we obviously have

$$\sum_{d|h} \mu(d) = 1.$$

Now we assert:

The cyclotomic polynomials are given by

$$\Phi_h(x) = \prod_{d|h} (x^d - 1)^{\mu\left(\frac{h}{d}\right)}. \tag{3}$$

In order to prove this, it suffices to show that the functions on the right-hand side satisfy equation (2), i.e., that

$$x^h - 1 = \prod_{d|h} \prod_{d'|d} (x^{d'} - 1)^{\mu\left(\frac{d}{d'}\right)}.$$

The exponents of a fixed $x^{d'} - 1$ are the numbers $\mu\left(\frac{d}{d'}\right)$, where d is a divisor of h and a multiple of d', i.e., they are all of the form $\mu(\lambda)$, where λ is any divisor of $\frac{h}{d'}$. As we have seen, the sum of these exponents is nearly always zero; only in the case $\frac{h}{d'} = 1$ has it the value 1. Therefore, from the entire double product on the right only one factor $x^h - 1$ is left with the exponent 1. Thus equation (2) is satisfied by the functions (3).

EXAMPLES:

$$\Phi_{12}(x) = (x^{12}-1)^{+1}(x^6-1)^{-1}(x^4-1)^{-1}(x^2-1)^{+1}$$
$$= (x^6+1)^{+1}(x^2+1)^{-1} = x^4 - x^2 + 1;$$
$$\Phi_{q^\nu}(x) = (x^{q^\nu}-1)^{+1}(x^{q^{\nu-1}}-1)^{-1}$$
$$= 1 + x^{q^{\nu-1}} + x^{2q^{\nu-1}} + \cdots + x^{(q-1)q^{\nu-1}}$$

for every prime number q.

The polynomial $\Phi_h(x)$ may very well be reducible; thus, for example, for characteristic 11 we have

$$\Phi_{12}(x) = x^4 - x^2 + 1 = (x^2 - 5x + 1)(x^2 + 5x + 1).$$

But later on (Section 53) we shall see that in the prime field of characteristic zero the polynomial $\Phi_h(x)$ is irreducible, which implies that all primitive h-th roots of unity are conjugate. We have already seen in Section 24 that this is true for all prime numbers h in virtue of the Eisenstein theorem; for $\Phi_8 = x^4 + 1$ and $\Phi_{12} = x^4 - x^2 + 1$ it was the content of Ex. 3, Section 24, and Ex. 5, Section 23.

A very useful theorem is the following:

If ζ is a h-th root of unity, we have

$$1 + \zeta + \zeta^2 + \cdots + \zeta^{h-1} = \begin{cases} h (\zeta = 1) \\ 0 (\zeta \neq 1) \end{cases}.$$

The proof is obtained at once from the summation formula of the geometric series: For $\zeta \neq 1$ we get

$$\frac{1-\zeta^h}{1-\zeta} = 0.$$

EXERCISES. 1. The field of the h-th roots of unity is at the same time the field of the $2h$-th roots of unity for odd h.

2. The fields of the third and fourth roots of unity over the field of rationals are quadratic. Express these roots of unity in terms of square roots.

3. The field of the eight roots of unity is quadratic with respect to the Gaussian number field $\Gamma(i)$. Express a primitive eighth root of unity by means of a square root of an element of $\Gamma(i)$.

4. In case of characteristic p the $(p^\lambda h)$-th roots of unity are at the same time h-th roots of unity. (This justifies the restriction $h \not\equiv 0 (p)$) imposed at the outset.)

5. The "*cyclotomic equation*" $\Phi_h(x) = 0$ is always normal.

37. • GALOIS FIELDS (FINITE COMMUTATIVE FIELDS)

We have met fields with a finite number of elements in the prime fields of characteristic p. Finite fields are known as *Galois fields* after their discoverer Galois. We shall first investigate their general properties.

Let Δ be a Galois field, and let q be the number of its elements.

The characteristic of Δ cannot be zero; for in this case the prime field Π in Δ would already have infinitely many elements. Let p be the characteristic. Then the prime field Π is 1-isomorphic with the integral residue class ring modulo p, and has p elements.

Since there are only a finite number of elements in Δ, there is in Δ a maximal set of linearly independent elements $\alpha_1, \ldots, \alpha_n$ with respect to Π. n is the degree of the field $(\Delta : \Pi)$, and every element of Δ is of the form

$$c_1\alpha_1 + \cdots + c_n\alpha_n \tag{1}$$

with uniquely determined coefficients c_i in Π.

For every coefficient c_i, p values are possible; thus there are exactly p^n expressions of form (1). Since they express the totality of the elements of the field, it follows that

$$q = p^n.$$

Thus we have proved:

The number of the elements of a Galois field is a power of the characteristic p; the exponent denotes the degree of the field $(\Delta : \Pi)$.

With the zero element left out, every skew field is a multiplicative group. In case of a Galois field the group is Abelian of order q -1. The order of an arbitrary element α must be a divisor of $q-1$, whence it follows that

$$\alpha^{q-1} = 1 \qquad \text{for every } \alpha \neq 0.$$

On multiplying this equation by α, we get another equation

$$\alpha^q - \alpha = 0,$$

which is also valid for $\alpha = 0$. Hence all field elements are roots of the polynomial $x^q - x$. If $\alpha_1, \ldots, \alpha_q$ are the field elements, $x^q - x$ must be divisible by

$$\prod_1^q (x - \alpha_i).$$

Since the degrees are equal, we have

$$x^q - x = \prod_1^q (x - \alpha_i).$$

Thus Δ arises from Π by the adjunction of all roots of a single polynomial $x^q - x$. Hence Δ is uniquely determined, except for 1-isomorphism (Section 35). Thus we see:

For given p and n all commutative fields with p^n elements are 1-isomorphic.

We shall now show that for every $n > 0$ and every p there actually exists a field with $q = p^n$ elements.

We start with the prime field Π of characteristic p, and form a field over Π in which $x^q - x$ completely resolves into linear factors. In this field we consider the set of the roots of $x^q - x$. This set is a field; for, according to Section 30, Ex. 1, $x^{p^n} = x$ and $y^{p^n} = y$ imply

$$(x-y)^{p^n} = x^{p^n} - y^{p^n},$$

and, provided $y \neq 0$,

$$\left(\frac{x}{y}\right)^{p^n} = \frac{x^{p^n}}{y^{p^n}},$$

according to which the difference and the quotient of two roots are again roots.

The polynomial $x^q - x$ has only simple roots; for its derivative is

$$q x^{q-1} - 1 = -1,$$

since $q \equiv 0 (p)$, and -1 never becomes zero. Thus the set of its roots is a field with q elements.

Thus we have proved:

For every power of a prime $q = p^n (n > 0)$ there exists one, and except for isomorphism only one, Galois field with precisely q elements. The elements are the roots of the polynomial $x^q - x$.

The Galois field with precisely p^n elements will be denoted by $GF(p^n)$.

We put $q - 1 = h$, and note that all nonzero elements of the Galois field are roots of $x^h - 1$ and therefore h-th roots of unity. Since h is relatively prime to p, everything said in the preceding section is valid for these roots of unity:

All field elements different from zero are powers of a single primitive h-th root of unity. Or: The multiplicative group of the Galois field is cyclic.

These theorems completely reveal the structure of finite commutative fields.

It is easy to determine all subfields of $GF(p^n)$. Every subfield is of degree m, a divisor of n, and therefore consists of p^m elements which are characterized by the fact that they must be roots of $x^{p^m} - x$. For every positive divisor m of n there actually exists such a subfield; for if m is a divisor of n, $p^m - 1$ is a divisor of $p^n - 1 = (p^m - 1)(p^{n-m} + p^{n-2m} + \cdots + p^m + 1)$ so that $x^{p^m-1} - 1$ is a divisor of $x^{p^n-1} - 1$, and $x^{p^m} - x$ is a divisor of $x^{p^n} - x$. Since the latter polynomial completely resolves in $GF(p^n)$, the former must resolve, too, and its roots form a $GF(p^m)$. Thus we have proved: *For every divisor $m > 0$ of n the field $GF(p^n)$ has one, and only one, subfield $GF(p^m)$ of degree m. An element $\alpha \neq 0$ belongs to the subfield if it satisfies the equation $\alpha^{p^m-1} = 1$, that is, if its order (in the multiplicative group) is a divisor of $p^m - 1$.*

In the next section we shall make use of the following theorem:

For every element a, a Galois field of characteristic p contains exactly one p-th root $a^{\frac{1}{p}}$.

PROOF. For every element x there exists a p-th power x^p in the field. Different elements have different p-th powers, since

$$x^p - y^p = (x - y)^p.$$

Therefore, there are exactly as many p-th powers in the field as there are elements. Thus all elements are p-th powers.

Finally we shall determine the 1-automorphisms of the field $\Sigma = GF(p^m)$.

First of all $\alpha \to \alpha^p$ is an automorphism. For, by the foregoing theorem, the mapping is one-to-one, and on the other hand we have

$$(\alpha + \beta)^p = \alpha^p + \beta^p,$$
$$(\alpha\beta)^p = \alpha^p \beta^p.$$

The powers of this automorphism carry α over into α^p, $\alpha^{p^2}, \ldots, \alpha^{p^m} = \alpha$. Thus we have found m automorphisms.

We shall see in Section 38 that there cannot be more than m 1-automorphisms so that the above determined m 1-automorphisms $\alpha \to \alpha^{p^\nu}$ are the *only ones*.

For the special case $n = 1$ the theorems valid for $GF(p^n)$, when applied to the residue class ring $C/(p)$, yield well known theorems of elementary number theory, namely:

1. The number of roots mod p of a congruence mod p is at most equal to the degree of the congruence.

2. Fermat's theorem

$$a^{p-1} \equiv 1 (p) \quad \text{for} \quad a \not\equiv 0 (p)$$

is a special case of the theorem valid for $GF(p^n)$

$$a^{p^n - 1} = 1 \quad \text{for} \quad a \neq 0.$$

3. There is a "primitive number ζ modulo p" such that every number b, relatively prime to p, is congruent modulo p to a power of ζ. (Or: The group of the residue classes mod p, excluding the null class, is cyclic.)

4. The product of all nonzero elements $a_1, a_2, \ldots, a_k$ of a $GF(p^n)$ is equal to -1, since

$$x^k - 1 = \prod_1^k (x - a_\nu).$$

For $n = 1$ we get "Wilson's theorem":

$$(p-1)! \equiv -1 (p).$$

EXERCISES. If α in $GF(p^n)$ is a root of the polynomial $f(x)$ of degree m irreducible in $\Pi[x]$, all roots (or elements conjugate to α) are given by

$$\alpha^p, \alpha^{p^2}, \ldots, \alpha^{p^m} = \alpha.$$

2. If r is relatively prime to $p^n - 1$, every element of $GF(p^n)$ is an r-th power. If r is divisor of $p^n - 1$, then those, and only those, elements α of $GF(p^n)$ are r-th powers which satisfy the equation

$$\alpha^{\frac{p^n - 1}{r}} = 1.$$

In elementary number theory, the numbers representing remainder classes satisfying this equation are called "r-th power remainders" (mod p) because their remainder classes (mod p) are r-th powers in the field of remainder classes (mod p). In case $r = 2$ we get the well-known "square residues" (mod p).

3. If a prime ideal $\mathfrak{p}$ in a commutative ring $\mathfrak{o}$ possesses only a finite number

of residue classes, $\mathfrak{o}/\mathfrak{p}$ is a Galois field.

4. Investigate the residue class rings modulo the prime ideals

$$(1+i),\ (3),\ (2+i),\ (7)$$

in the ring of the Gaussian integers.

5. Write the equation irreducible in $GF(3)$ for a primitive eighth root of unity in $GF(9)$, and also the equation irreducible in $GF(2)$ for a primitive seventh root of unity in $GF(8)$.

6. For every p and m there are polynomials $f(x)$ of degree m which are irreducible mod p. They are all divisors of $x^{p^m} - x \pmod{p}$.

An interesting property of the Galois fields was proved by C. Chevalley: *Abh. math. Sem. Hamburg*, Vol. II (1935), p. 73.

38. SEPARABLE AND INSEPARABLE EXTENSIONS

Again let Δ be a commutative field.

We ask: Can a polynomial irreducible in $\Delta[x]$ have multiple roots in an extension field?

For $f(x)$ to possess multiple zeros, it is necessary that $f(x)$ and $f'(x)$ have a non-constant factor in common which, by Section 18, can already be computed in $\Delta[x]$. If $f(x)$ is irreducible, it cannot have a non-constant factor in common with a polynomial of lower degree, and, therefore, we must have $f'(x) = 0$.

We put

$$f(x) = \sum_0^n a_\nu x^\nu,$$

$$f'(x) = \sum_1^n \nu a_\nu x^{\nu-1}.$$

If $f'(x)$ is to be equal to 0, every coefficient has to vanish:

$$\nu a_\nu = 0 \qquad (\nu = 1, 2, \ldots).$$

In case of characteristic zero it follows that $a_\nu = 0$ for all $\nu \neq 0$. Thus a non-constant polynomial cannot have a multiple root. In case of characteristic p, $\nu a_\nu = 0$ is possible even for $a_\nu \neq 0$, but then we must have

$$\nu \equiv 0 (p).$$

Hence, if $f(x)$ has a multiple root, all terms have to vanish, except the terms $a_\nu x^\nu$ with $\nu \equiv 0(p)$ so that $f(x)$ is of the form

$$f(x) = a_0 + a_p x^p + a_{2p} x^{2p} + \cdots$$

If, conversely, $f(x)$ is of this form, $f'(x) = 0$. In this case we may write:

$$f(x) = \varphi(x^p).$$

Thus we have proved: *For characteristic zero a polynomial $f(x)$ irreducible in $\Delta[x]$ has only simple roots, and for characteristic p the polynomial $f(x)$ (provided it is non-constant) has multiple zeros only if $f(x)$ can be written as a function of x^p.*

It may be that, in the latter case, $\varphi(x)$ is itself a function of x^p. Then $f(x)$ is a function of x^{p^2}. Let $f(x)$ be a function of x^{p^e},

$$f(x) = \psi\left(x^{p^e}\right),$$

but not a function of $x^{p^{e+1}}$. Then, of course, $\psi(y)$ is irreducible. Moreover, $\psi'(y) \neq 0$; otherwise we would have $\psi(y) = \chi(y^p)$, which would imply $f(x) = \chi\left(x^{p^{e+1}}\right)$, contrary to hypothesis.—Therefore, $\psi(y)$ has only simple zeros.

We now resolve $\psi(y)$ into linear factors [9] in an extension field:

$$\psi(y) = \prod_1^{n_0} (y - \beta_i).$$

It follows that

$$f(x) = \prod_1^{n_0} \left(x^{p^e} - \beta_i\right).$$

Let α_i be a root of $x^{p^e} - \beta_i$. Then we have

$$\alpha_i^{p^e} = \beta_i,$$

$$x^{p^e} - \beta_i = x^{p^e} - \alpha_i^{p^e} = (x - \alpha_i)^{p^e}.$$

Therefore, α_i is a root of $x^{p^e} - \beta_i$, with multiplicity p^e and we have

$$f(x) = \prod_1^{n_0} (x - \alpha_i)^{p^e}.$$

Thus all roots of $f(x)$ have the same multiplicity p^e.

The degree n_0 of the polynomial ψ is called the *reduced degree* of $f(x)$ (or of α_i); e is called the *exponent* of $f(x)$ (or of α_i) with respect to Δ. Between the degree, the reduced degree, and the exponent the following relationship exists:

$$n = n_0 p^e.$$

n_0, at the same time, is the number of different roots of $f(x)$.

If ϑ is a root of a polynomial irreducible in $\Delta[x]$ which has only separated (simple) roots, then ϑ is called *separable* or *of the first kind* [10] with respect to Δ. The irreducible polynomial $f(x)$, having separate roots, is also called *separable*. In the opposite case the algebraic element ϑ and the irreducible polynomial $f(x)$ are called *inseparable* or *of the second kind*. Finally, if all elements of an algebraic extension field Σ are separable with respect to Δ, the extension is called *separable* with respect to Δ, and any other algebraic extension field is called *in-*

[9] Without loss of generality the greatest coefficient of $\psi(y)$ may be equated to 1: n_0 is the degree of $\psi(y)$.

[10] The expression "of the first kind" (*von erster Art*) is due to Steinitz. I suggested the term "separable."

separable.

If the characteristic is zero, every irreducible polynomial is separable (and so is every algebraic extension field); for characteristic p only the polynomials with exponent $e = 0$ (and therefore with reduced degree $n_0 = n$) are separable. In case of characteristic p an irreducible non-constant polynomial $\varphi(x)$ is inseparable if, and only if, it may be written as a polynomial in x^p.

We shall see later that most of the important and interesting field extensions are separable, and that there are numerous classes of fields which are not capable of any inseparable extension (the so-called "perfect fields"). It is for this reason that in the sequel all investigations dealing with inseparable extensions are printed in small type.

We now consider the algebraic field $\Sigma = \Delta(\vartheta)$. While the degree n of the defining equation $f(x) = 0$ denotes at the same time the degree of the field $(\Sigma : \Delta)$, the reduced degree n_0 also denotes the *number of isomorphisms* of the field Σ in the following more precisely defined sense: We consider only such isomorphisms $\Sigma \cong \Sigma'$ as leave all elements of the subfield Σ fixed so that they carry Σ into equivalent fields Σ' (*"relative isomorphisms of Σ with respect to Δ"*), and we assume that Σ' as well as Σ lie within a suitably chosen extensions field Ω. For these isomorphisms we have the theorem:

If the extension field Ω is suitably chosen, $\Sigma = \Delta(\vartheta)$ has precisely n_0 relative isomorphisms. There is no Ω in which Σ has more than n_0 such isomorphisms.

PROOF. Every relative isomorphism must carry ϑ into a conjugate element ϑ' in Ω [root of the same irreducible equation $f(x) = 0$]. If we choose Ω so that $f(x)$ completely resolves into linear factors in Ω, then ϑ actually has n_0 conjugates $\vartheta, \vartheta', \ldots,$ and the fields $\Delta(\vartheta), \Delta(\vartheta'), \ldots$ are indeed conjugate or equivalent. But no matter what Ω we chose, ϑ never has more than n_0 conjugates. Now we note that, by giving $\vartheta \to \vartheta'$, a relative isomorphism $\Delta(\vartheta) \cong \Delta(\vartheta')$ is completely determined; for if ϑ is to go into ϑ', and if every element in Δ is to remain fixed, then

$$\Sigma a_k \vartheta^k \qquad (a_k \in \Delta)$$

must go into

$$\Sigma a_k \vartheta'^k,$$

and this determines the isomorphism.

If, in particular, ϑ is separable, we have $n_0 = n$, and therefore the number of relative isomorphisms is equal to the degree of the field.

When, in the following, we speak of the (relative) isomorphisms of $\Sigma = \Delta(\vartheta)$, of the conjugates to ϑ, or of the conjugate fields to Σ (with respect to Δ) we always mean, respectively, the isomorphisms or conjugates in a suitably chosen field Ω for which we can, as above, always take the decomposition field of $f(x)$, i.e., the

smallest field which is normal with respect to Δ and includes Σ.

If we have a fixed extension field in which *every* equation $f(x) = 0$ completely resolves into linear factors (such as in the field of complex numbers), then we may, once and for all, take this fixed extension field for Ω and always omit the addendum "in Ω" in statements regarding isomorphisms. This is standard practice, for example in the theory of number fields. In Section 62 we shall see that such an Ω is always available.

EXERCISES. 1. If Π is a field of characteristic p and x an indeterminate, then the equation $z^p - x = 0$ is irreducible in $\Pi(x)[z]$, and the field $\Pi\left(x^{\frac{1}{p}}\right)$ defined by this equation is inseparable over $\Pi(x)$.

2. Construct the relative isomorphisms with respect to the rational field Γ:

a) of the field of the fifth roots of unity,

b) of the field $\Gamma\left(\sqrt[3]{2}\right)$.

3. If $\vartheta^{p^e} = \gamma$ lies in Δ, but if $\vartheta^{p^{e-1}}$ does not, the polynomial $x^{p^e} - \gamma$ is irreducible in $\Delta[x]$.

A generalization of the above theorem is the following:

If an extension field Σ arises from Δ by the successive adjunction of m algebraic elements $\alpha_1, \ldots, \alpha_m$, and if every α_i is the root of an equation irreducible in $\Delta(\alpha_1, \ldots, \alpha_{i-1})$ and of reduced degree n_i', then, in a suitable extension field Ω, Σ has exactly $\prod_1^m n_i'$ relative isomorphisms with respect to Δ, and in no extension field are there more than $\prod_1^m n_i'$ such isomorphisms of Σ.

PROOF. The theorem has just been proved for $m = 1$. Let its correctness be assumed for $\Sigma_1 = \Delta(\alpha_1, \ldots, \alpha_{m-1})$; let there be no more than just $\prod_1^{m-1} n_i'$ relative isomorphisms of Σ_1 in a suitable Ω_1. Let one of these $\prod_1^{m-1} n_i'$ isomorphisms be $\Sigma_1 \to \bar{\Sigma}_1$. We now assert that, in a suitable Ω, this isomorphism may be continued to an isomorphism $\Sigma = \Sigma_1(\alpha_m) \cong \bar{\Sigma} = \bar{\Sigma}_1(\bar{\alpha}_m)$ in no more than precisely n_m' ways.

α_m satisfies in Σ_1 an equation $f_1(x) = 0$ with exactly n_m' distinct roots. Let $f_1(x)$ go into $\bar{f}_1(x)$ under the isomorphism $\Sigma \to \bar{\Sigma}_1$. Then, in a suitable extension field, $f_1(x)$ has again n_m' distinct roots and no more. Let one of these roots be $\bar{\alpha}_m$. With $\bar{\alpha}_m$ chosen, the isomorphism $\Sigma_1 \cong \bar{\Sigma}_1$ can be continued in one, and only one, way to an isomorphism $\Sigma_1(\alpha_m) \cong \bar{\Sigma}_1(\bar{\alpha}_m)$ with $\alpha_m \to \bar{\alpha}_m$; this continuation is given by the formula

$$\sum c_k \alpha_m^k \to \sum \bar{c}_k \bar{\alpha}_m^k.$$

Since we may choose α_m in n_m' ways, there are n_m' such continuations for every isomorphism $\Sigma_1 \to \bar{\Sigma}_1$ chosen. Since this isomorphism may itself be chosen in $\prod_1^{m-1} n_i'$ ways, there are altogether (in such an extension field Ω in which all equations under consideration are completely decomposed)

$$\prod_1^{m-1} n_i' \cdot n_m' = \prod_1^m n_i'$$

and no more relative isomorphisms. Thus, by complete induction, we get the required result.

If n_i is the full (non-reduced) degree of α_i with respect to $\Delta(\alpha_1, \ldots, \alpha_{i-1})$, then n_i is at the same time the degree of $\Delta(\alpha_1, \ldots, \alpha_i)$ with respect to $\Delta(\alpha_1, \ldots, \alpha_{i-1})$; thus the degree of the field $(\Sigma : \Delta)$ is equal to $\prod_1^m n_i$. If we compare this number with the number of the isomorphisms $\prod_1^m n_i'$, it follows:

The number of relative isomorphisms of a finite extension field $\Sigma = \Delta(\alpha_1, \ldots, \alpha_m)$ with respect to Δ (in a suitable extension field Ω) is equal to the degree of the field $(\Sigma : \Delta)$, if every α_i is separable with respect to the corresponding $\Delta(\alpha_1, \ldots, \alpha_{i-1})$. If, on the other hand, there is one α_i which is inseparable, the number of isomorphisms is smaller than the degree of the field.

This theorem immediately yields a number of important deductions. In the first place, the theorem states that the property, that every α_i is separable with respect to the field of the preceding ones, is a property of the field Σ, regardless of the choice of the generator α_i. Since any element β of the field may be chosen as the first generator, it follows at once that every element β of the field Σ is separable, provided that all α_i are separable in the sense stated. Therefore we have the following theorem:

If we successively adjoin elements $\alpha_1, \ldots, \alpha_m$ to Δ, and if every α_i is separable with respect to the field of the preceding ones, the resulting field

$$\Sigma = \Delta(\alpha_1, \ldots, \alpha_m)$$

will be separable over Δ.

In particular: *Sum, difference, product, and quotient of separable elements are separable.*

Furthermore: *If β is separable with respect to Σ, and Σ separable with respect to Δ, then β is separable with respect to Δ.* For β satisfies an equation with a finite number of coefficients $\alpha_1, \ldots, \alpha_m$ in Σ so that it is separable with respect to $\Delta(\alpha_1, \ldots, \alpha_m)$; hence

$$\Delta(\alpha_1, \ldots, \alpha_m, \beta)$$

is also separable.

Finally we have: *The number of relative isomorphisms of a separable finite extension field Σ of Δ is equal to the degree of the field $(\Sigma : \Delta)$.*

Since, by the foregoing, all rational operations performed on separable elements yield again separable elements, the separable elements themselves form a field Ω_0 in an arbitrary extension field Ω of Δ. Ω_0 may also be called the greatest separable extension of Δ within Ω.

If Ω is algebraic with respect to Δ, but not necessarily separable, then the p^e-th power of every element α of Ω lies in Ω_0 if e is the exponent of this element; for from the considerations made at the beginning of this section it follows immediately that α^{p^e} satisfies an equation with distinct roots only. Thus we have:

Ω arises from Ω_0 by extracting only p^e-th roots.

If, in particular, Ω is finite with respect to Δ, the exponents e are of course limited. The largest among them, which we shall again denote by e, is called the *exponent* of Ω. The degree of Ω_0 is called the *reduced degree* of Ω.

Of course, the extraction of the p^e-th roots can also be effected by successive extraction of p-th roots. When we extract a p-th root which is not yet in the field (i.e., on adjoining a root of an irreducible equation $x^p - \beta = 0$), the degree of the field multiplies by p. Thus, after having adjoined f such p-th roots, we finally have

$$(\Omega : \Delta) = (\Omega_0 : \Delta) \cdot p^f \quad \text{or}$$

$$\text{degree} = \text{reduced degree} \cdot p^f,$$

just as in simple inseparable extensions.

p^e-th roots obey very simple rules. If α is a p^e-th root in β, then

$$x^{p^e} - \beta = x^{p^e} - \alpha^{p^e} = (x - \alpha)^{p^e},$$

as we have already seen. Therefore, a p^e-th root in β is *uniquely determined* in any field in which it exists. Moreover, we have

$$\sqrt[p^e]{\alpha+\beta}=\sqrt[p^e]{\alpha}+\sqrt[p^e]{\beta},$$

$$\sqrt[p^e]{\alpha\beta}=\sqrt[p^e]{\alpha}\sqrt[p^e]{\beta},$$

as can be seen by raising these expressions to the p-th power.

EXERCISE. 4. If, for a finite inseparable extension, e and f are defined as above, we have $e \leqq f$. In a simple extension $e=f$. Give an example for $e<f$. [Adjunction of the p-th roots from two or more indeterminates.]

39. PERFECT AND IMPERFECT FIELDS. ROOT FIELDS

A field Δ is called *perfect* if every polynomial $f(x)$ irreducible in $\Delta[x]$ is separable. Any other field is called *imperfect.*

The question as to when a field is perfect is answered in the following two theorems:

I. *Fields of characteristic zero are always perfect.*

PROOF. See Section 38.

II. *A field of characteristic p is perfect if, and only if, there exists within the field itself a p-th root for every element.*

PROOF. If, for every element, there exists a p-th root in the field, then every polynomial $f(x)$ containing only powers of x^p is a p-th power, since

$$f(x)=\sum_k a_k(x^p)^k=\Big\{\sum_k \sqrt[p]{a_k}\,x^k\Big\}^p;$$

hence every irreducible polynomial is separable in this case, so the field is perfect.

If, on the other hand, there exists in the field an element α which is not a p-th power, we consider the polynomial

$$f(x)=x^p-\alpha.$$

Let $\varphi(x)$ be an irreducible factor of $f(x)$. After the adjunction of $\sqrt[p]{\alpha}=\beta$, $f(x)$ resolves into linear factors $(x-\beta)$ which are all equal, and so $\varphi(x)$, as divisor of $f(x)$, is likewise a power of $(x-\beta)$. If $\varphi(x)$ were linear, so that we would have $\varphi(x)=x-\beta$, then β would belong to the field Δ, contrary to hypothesis. Hence $\varphi(x)=(x-\beta)^k$, where $k>1$ is an inseparable irreducible polynomial over Δ; consequently, Δ is an imperfect field. Incidentally, by Section 38, the degree of $\varphi(x)$ is necessarily divisible by p; in this case it is equal to p, i.e., $\varphi(x)=f(x)$.

From II and a theorem in Section 37 we infer immediately:

All Galois fields are perfect fields.

In an algebraically closed field (for the definition see Section 62) every polynomial is linear; hence

All algebraically closed [11] *fields are perfect fields.*

[11] Translator's note: Also called "algebraically complete" fields.

The next two theorems follow immediately from the definition of a perfect field:

Every algebraic extension of a perfect field is separable with respect to the latter.

To an imperfect field there exist inseparable extensions.

These inseparable extensions are obtained by adjoining any root of a prime inseparable polynomial.

In the proof of II we remarked that in a perfect field of characteristic p every polynomial $f(x)$ depending solely on x^p is a p-th power. The proof shows that this statement also holds for polynomials in several variables $f(x, y, z, \ldots)$ which can be written as polynomials in $x^p, y^p, z^p, \ldots$. This is another useful property of perfect fields of characteristic p.

ROOT FIELDS. Let Δ be any field of characteristic p. If, with every element x of Δ, we associate its p-th power x^p, we obtain a correspondence of Δ to a subset which we shall call Δ^p. In case of a perfect Δ we have $\Delta^p = \Delta$, as we saw. In any event, however, the correspondence is one-to-one; for since

$$a^p - b^p = (a - b)^p,$$

$a^p = b^p$ necessarily entails $a = b$. Moreover, the correspondence is an isomorphism since

$$a^p + b^p = (a + b)^p,$$
$$a^p \cdot b^p = (a \cdot b)^p.$$

Hence Δ^p is a field isomorphic with Δ.

Conversely, we can in exactly the same manner construct from a field Δ an isomorphic extension field $\Delta^{\frac{1}{p}}$ whose p-th power is Δ. All that is necessary is to adjoin all p-th roots $a^{\frac{1}{p}}$ of the elements of Δ, insofar as they do not lie in Δ, and to impose on them the following rules of operation:

$$a^{\frac{1}{p}} + b^{\frac{1}{p}} = (a + b)^{\frac{1}{p}},$$
$$a^{\frac{1}{p}} \cdot b^{\frac{1}{p}} = (a \cdot b)^{\frac{1}{p}}.$$

The field properties for $\Delta^{\frac{1}{p}}$ can be proved without difficulty with the aid of the isomorphism $a \to a^{\frac{1}{p}}$; the proof may be left to the reader. Except for equivalent extensions, $\Delta^{\frac{1}{p}}$ is uniquely determined by Δ.

If Δ is perfect, we have of course $\Delta^{\frac{1}{p}} = \Delta$ and vice versa.

We call $\Delta^{\frac{1}{p}}$ the *root field* of Δ. If we construct the roof field again for $\Delta^{\frac{1}{p}}$, etc., we obtain a sequence of fields

$$\Delta, \Delta^{\frac{1}{p}}, \Delta^{\frac{1}{p^2}}, \ldots$$

The union of these fields is clearly a perfect field; it is the smallest perfect field which includes Δ.

EXERCISES. 1. Give the proofs.

2. Every algebraic extension of a perfect field is perfect.

3. Every finite algebraic extension of an imperfect field is imperfect.

4. Construct the root field for $\Pi(x)$, where Π is a perfect field of characteristic p and x an indeterminate; similarly, the smallest comprehending perfect field.

40. SIMPLICITY OF ALGEBRAIC EXTENSIONS. THEOREM OF THE PRIMITIVE ELEMENT

We shall now investigate the conditions under which a commutative finite extension Σ of a field Δ is simple, i.e., the conditions under which it can be obtained by the adjunction of a single generating or "*primitive*" element. This question is answered for numerous cases by the following *Theorem on the Primitive Element*:

Let $\Delta(\alpha_1, \ldots, \alpha_h)$ *be a finite algebraic extension field of* Δ, *and let* $\alpha_2, \ldots, \alpha_h$ *be separable elements.*[12] *Then* $\Delta(\alpha_1, \ldots, \alpha_h)$ *is a simple extension*:

$$\Delta(\alpha_1, \ldots, \alpha_h) = \Delta(\vartheta).$$

PROOF. First, we prove the theorem for two elements α, β, one of which, say β, shall be separable. Let $f(x) = 0$ be the irreducible equation for α, and $g(x) = 0$ that for β. We take a field in which $f(x)$ and $g(x)$ resolve completely. Let the distinct zeros of $f(x)$ *be* $\alpha_1, \ldots, \alpha_r$, and let those of $g(x)$ be $\beta_1, \ldots, \beta_s$; let, for example, $\alpha_1 = \alpha$, $\beta_1 = \beta$.

We may assume that Δ has an infinite number of elements. For if Δ were finite, $\Delta(\alpha, \beta)$ would be, too. For finite fields the existence of a primitive element (even of a primitive root of unity, all field elements of which, except the zero, are powers) was already proved in Section 37.

For $k \neq 1$ we have $\beta_k \neq \beta_1$ so that the equation

$$\alpha_i + x\beta_k = \alpha_1 + x\beta_1$$

has at most one root x in Δ for every i and every $k \neq 1$. If we take c different from the roots of all these linear equations, we have

$$\alpha_i + c\beta_k \neq \alpha_1 + c\beta_1$$

for every i and $k \neq 1$.

We let

$$\vartheta = \alpha_1 + c\beta_1 = \alpha + c\beta.$$

Then ϑ is an element of $\Delta(\alpha, \beta)$. I assert that ϑ already has the property of the required primitive element: $\Delta(\alpha, \beta) = \Delta(\vartheta)$.

The element β satisfies the equations

$$g(\beta) = 0,$$
$$f(\vartheta - c\beta) = f(\alpha) = 0,$$

with coefficients in $\Delta(\vartheta)$. The polynomials $g(x)$, $f(\vartheta - cx)$ have only the root β in common; for we have for the other roots $\beta_k (k \neq 1)$ of the first equation

$$\vartheta - c\beta_k \neq \alpha_i \qquad (i = 1, \ldots, r),$$

and so

$$f(\vartheta - c\beta_k) \neq 0.$$

[12] It is immaterial whether α_1, and so the entire field, is separable.

β is a simple root of $g(x)$; therefore, $g(x)$ and $f(\vartheta - cx)$ have but one linear factor $x-\beta$ in common. The coefficients of this greatest common divisor must lie in $\Delta(\vartheta)$ already; thus β lies in $\Delta(\vartheta)$. From $\alpha=\vartheta-c\beta$ the same thing follows for α, so that we have indeed $\Delta(\alpha,\beta)=\Delta(\vartheta)$.

This completes the proof of our theorem for $h=2$. Once it is proved for $h-1(\geqq 2)$, we have

$$\Delta(\alpha_1,\ldots,\alpha_{h-1})=\Delta(\eta),$$

and so

$$\Delta(\alpha_1,\ldots,\alpha_h)=\Delta(\eta,\alpha_h)=\Delta(\vartheta),$$

according to the already proved part of the theorem; thus it follows that the theorem holds for h.

Conclusion. *Every separable finite extension is simple.*

This theorem greatly simplifies the investigation of the finite separable extensions, since we easily master the structure and isomorphisms of these extensions by means of the very simple basis representation

$$\sum_0^{n-1} a_k \vartheta^k .$$

For example, we now have a new proof of the fact proved in Section 38 (small type) by means of successive continuation of isomorphisms, that *the number of relative isomorphisms with respect to Δ of a finite separable extension Σ of Δ is equal to the degree $(\Sigma:\Delta)$*. For simple separable extensions this statement was already proved in Section 38, and every finite separable extensions is, as we now know, a simple extension.[13]

In case of characteristic zero every finite extension is separable and therefore simple. But even in case of characteristic p we can tell exactly when a finite extension is simple:

A finite extension Σ of a field Δ of characteristic p is simple when, and only when,

$$n=n_0 p^e, \tag{1}$$

where n is the degree, n_0 the reduced degree, and e the exponent of the extension.

PROOF. First a preliminary remark: If $\Sigma=\Delta(\alpha_1,\ldots,\alpha_r)$, the exponent of Σ is equal to the maximum e of the exponents of $\alpha_1,\ldots,\alpha_r$. For, first of all, the exponent of Σ is surely $\geqq e$. On the other hand, however, since all elements of Σ can be expressed rationally in terms of $\alpha_1,\ldots,\alpha_r$ (with coefficients in Δ), the powers of the elements of Σ may be expressed rationally in terms of $\alpha_1^{p^e},\ldots,\alpha_r^{p^e}$. These p^e-th powers are separable elements; hence the exponent of Σ is exactly e. According to Section 32, the separable elements of Σ form a separable field Σ_0 in Σ.

Now let Σ be simple: $\Sigma=\Delta(\vartheta)$. Then $\vartheta^{p^e}\in\Sigma_0$. Thus, if we put $(\Sigma:\Sigma_0)=p^f$, then $f\leqq e$ and, as always, $e\leqq f$ so that $e=f$. From this and from $n=n_0 p^f$ follows relation (1).

---

[13] The earlier (lengthier) proof by means of successive continuation was more instructive; for the entire theory of inseparable extensions could be developed from it. To those readers who are mainly interested in separable extensions which, as a matter of fact, are the most important ones and occur most frequently, we recommend the above proof by means of the theorem of the primitive element.

Let, conversely, (1) be fulfilled. Σ_0 is a finite separable extension and therefore simple:

$$\Sigma_0 = \Delta(\alpha).$$

Select an element β of Σ which has exactly the exponent e. Then

$$\beta^{p^e} = \mu_0 \in \Sigma_0,$$

and $x^{p^e} - \mu_0$ is irreducible in Σ_0 since, otherwise, $\beta^{p^{e-1}}$ would have to lie in Σ_0 already (Section 38, ex. 3). Thus $\Sigma_0(\beta)$ is of degree p^e with respect to Σ_0, and of degree $n = n_0 p^e$ with respect to Δ. As Σ is also of degree n with respect to Δ, it follows that $\Sigma_0(\beta) = \Sigma$ or $\Sigma = \Delta(\alpha, \beta)$. Since α is separable, it follows from the theorem on the primitive element that Σ is simple. Q.E.D.

EXERCISES. 1. If x and y are indeterminates, the extension $\Delta\left(x^{\frac{1}{p}}, y^{\frac{1}{p}}\right)$ of $\Delta(x, y)$ is no longer simple.

2. For a characteristic $\neq 2$ and indeterminates x and y,

$$\Delta(\sqrt{x}, \sqrt{y}) = \Delta(x, y, \sqrt{x} + \sqrt{y})$$

is always valid, but not for characteristic 2.

41. NORMS AND TRACES

First, let Σ be a finite commutative extension field of Δ. The relative isomorphisms of Σ, which carry Σ into its conjugate fields, carry every element η of Σ into elements conjugate to η. Let us suppose for the sake of simplicity that Σ is a simple extension: $\Sigma = \Delta(\vartheta)$. Then we have

$$\eta = \psi(\vartheta) = a_0 + a_1\vartheta + \cdots + a_{n-1}\vartheta^{n-1}.$$

To perform the isomorphisms, we replace ϑ by its conjugates. Thus we obtain:

$$\eta_\nu = \psi(\vartheta_\nu) = a_0 + a_1\vartheta_\nu + \cdots + a_{n-1}\vartheta_\nu^{n-1}.$$

Here the conjugates ϑ_ν are numbered from 1 to n, and each one is counted as often as it occurs in the decomposition into linear factors of the polynomial $\varphi(t)$ irreducible in $\Delta[t]$ with ϑ being a root. Thus, for a separable ϑ every isomorphism is counted once, and for an inseparable ϑ it is counted several times.

We now turn to the elementary symmetric functions of $\eta_1, \ldots, \eta_n$. Among these the sum is known as the *trace of* η,

$$\sum_{\nu=1}^{n} \eta_\nu = S(\eta),$$

and the product is called the *norm of* η,

$$\prod_{\nu=1}^{n} \eta_\nu = N(\eta).$$

We extend the forming of traces and norms to polynomials with coefficients in Σ, making the convention that the isomorphisms $\vartheta \to \vartheta_\nu$ shall apply to the coefficients of these polynomials only, while the indeterminates shall remain untouched. For example, we have

$$N(z - \eta) = \prod_{\nu=1}^{n} (z - \eta_\nu). \tag{1}$$

The coefficients of this polynomial $G(z) = N(z-\eta)$ are exactly the elementary symmetric functions of $\eta_1, \ldots, \eta_n$. Thus we see that the norm concept is completely sufficient; the trace as well as the other elementary symmetric functions may be defined by means of the norm $N(z-\eta)$.

The above definition of norm, customary in the theory of algebraic numbers, has the advantage that the two fundamental properties of norms and traces, namely

$$\text{(2)} \qquad \begin{cases} N(\alpha\beta) = N(\alpha)\,N(\beta) \\ S(\alpha+\beta) = S(\alpha) + S(\beta) \end{cases}$$

are evident. The disadvantage of this definition, however, is the fact that it cannot be readily extended to non-simple field extensions, and it cannot be applied at all to skew fields and other hypercomplex systems. We shall, therefore, give some other definitions of the norm which can readily be generalized.

The polynomial (1) becomes, if $\eta_\nu = \psi(\vartheta_\nu)$ is substituted in it, a symmetric function of the roots ϑ_ν. Hence it can be expressed as a rational integral function in terms of the elementary symmetric functions of the ϑ_ν, i.e., in terms of the coefficients of $\varphi(t)$. The roots of $G(z) = N(z-\eta)$ are the elements η_ν conjugate to η. If $g(z)$ is the irreducible polynomial with η_1 as a root, $G(z)$ has the root η_1 in common with $g(z)$; so it is divisible by $g(z)$. Let $g(z)^r$ be the highest power of $g(z)$ which divides $G(z)$. Then

$$G(z) = g(z)^r\,h(z).$$

The second factor $h(z)$ has linear factors only of the form $z-\eta_\nu$, which divide $g(z)$. If $h(z)$ were no constant, $h(z)$ would have a root η_ν in common with $g(z)$ and thus would be divisible by $g(z)$. Hence $h(z)$ is a constant. If $g(z)$ is, like $h(z)$, so normed that the leading coefficient $= 1$, it follows that

$$G(z) = g(z)^r.$$

If we put

$$g(z) = z^m + b_1 z^{m-1} + \cdots + b_m,$$

we have

$$G(z) = g(z)^r = z^{mr} + r b_1 z^{mr-1} + \cdots + b_m^r$$

so that

$$\text{(3)} \qquad \begin{cases} S(\eta) = -r b_1 \\ N(\eta) = (-1)^{mr} b_m^r. \end{cases}$$

The degree of $G(z)$ is $n = mr$, so we have

$$\text{(4)} \qquad r = \frac{n}{m} = \frac{(\Sigma:\Delta)}{(\Delta(\eta):\Delta)} = (\Sigma:\Delta(\eta)).$$

The formulae (3) and (4) contain a trace and norm definition which remains meaningful when arbitrary non-commutative fields of finite rank over Δ are involved, provided Δ lies in the center of the skew field. However, in this

definition the properties (2) are not evident. Therefore, we lay down a third equivalent norm definition, which has the additional advantage that it can be extended to arbitrary hypercomplex systems.

Let $(\omega_1, \ldots, \omega_n)$ be a basis of the field Σ over Δ. We express all products $\eta\omega_j$ linearly in terms of this basis, viz.

$$\eta\omega_j = \sum_k c_{jk}\omega_k \tag{5}$$

and form the determinant D of the coefficient c_{jk}. We now assert that $D = N(\eta)$.

First we show that D is independent of the choice of the basis. If $(\omega_1', \ldots, \omega_n')$ is some other basis and

$$\omega_i' = \sum_j a_{ij}\omega_j\ , \tag{6}$$

$$\omega_k = \sum_l b_{kl}\omega_l'\ , \tag{7}$$

we find from (5), (6), and (7)

$$\eta\omega_i' = \sum_j \sum_k \sum_l a_{ij}c_{jk}b_{kl}\omega_l';$$

so the new coefficients are given by

$$c_{il}' = \sum_j \sum_k a_{ij}c_{jk}b_{kl}\,.$$

By the multiplication theorem of determinants, equation

$$D' = ABD \tag{8}$$

follows for the determinants D', A, B of the elements c_{ij}', a_{ij}, b_{kl}. If, in particular, we take $\eta = 1$, the c_{jk} and the c_{jk}' become equal to 1 for $j = k$. and equal to 0 for $j \neq k$ so that $D = D' = 1$; as a special case of (8) it follows that

$$1 = AB. \tag{9}$$

(8) and (9) imply $D' = D$: the determinant D is in fact independent of the choice of the basis.

We proceed to investigate the behavior of the determinant D in an extension of the field Σ. Let Ω be an extension field of Σ, let $(\sigma_1, \ldots, \sigma_n)$ be a basis for Σ over Δ, and $(\omega_1, \ldots, \omega_s)$ a basis for Ω over Σ so that $(\sigma_1\omega_1, \sigma_2\omega_1, \ldots \sigma_n\omega_1, \ldots, \sigma_1\omega_s, \ldots, \sigma_n\omega_s)$ is a basis for Ω over Δ. The determinant D_Σ of an element η of Σ is obtained from

$$\eta\sigma_j = \sum_k c_{jk}\sigma_k\,. \tag{10}$$

Upon multiplication of these formulae by ω_l, it follows that

$$\eta\sigma_j\omega_l = \sum_k c_{jk}\sigma_k\omega_l\,.$$

Thus the determinant of the coefficients D_Ω of the right side of (10) consists of s equal rectangular arrays (c_{jk}):

$$D = \begin{vmatrix} \boxed{\begin{matrix} c_{11} & \dots & c_{1n} \\ \dots & \dots & \dots \\ c_{n1} & \dots & c_{nn} \end{matrix}} & & 0 & \\ & \boxed{\begin{matrix} c_{11} & \dots & c_{1n} \\ \dots & \dots & \dots \\ c_{n1} & \dots & c_{nn} \end{matrix}} & & \\ 0 & & \ddots & \\ & & & \end{vmatrix}$$

Thus the determinant of η formed in the field Ω is

$$D_\Omega = D_\Sigma^s; \quad s = (\Omega : \Sigma). \tag{11}$$

For calculating the determinant D_Σ of an element η we may, by formula (11), limit ourselves to the smallest field which contains this element, i.e., to the field $\Delta(\eta)$ with the basis $(1, \eta, \eta^2, \dots, \eta^{m-1})$. Taking into account the defining equation $g(\eta) = 0$, we have with respect to this basis

$$\begin{aligned} \eta \cdot 1 &= \eta \\ \eta \cdot \eta &= \eta^2 \\ \eta \cdot \eta^2 &= \eta^3 \\ &\dots\dots\dots \\ \eta \cdot \eta^{m-1} &= \eta^m = -b_m - b_{m-1}\eta - \dots - b_1 \eta^{m-1}. \end{aligned}$$

The determinant of η in the field $\Delta(\eta)$ thus becomes

$$D_{\Delta(\eta)} = \begin{vmatrix} 0 & 1 & 0 \dots\dots 0 \\ 0 & 0 & 1 \dots\dots 0 \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ 0 & 0 & 0 \dots\dots 1 \\ -b_m & -b_{m-1} & -b_{m-2} \dots -b_1 \end{vmatrix} = (-1)^m b_m .$$

From (11) we now find

(12) $$D_\Sigma = D^r_{\Delta(\eta)} = (-1)^{mr} b^r_m, \qquad r = (\Sigma : \Delta(\eta)).$$

Comparing (3) and (4) with (12), we see that D_Σ is in fact the norm of η in the field Σ:

(13) $$N(\eta) = D_\Sigma.$$

If η is not a field element, but a polynomial in $\Sigma[x_1, x_2, \ldots]$, the same considerations are valid, except that the indeterminates $x_1, x_2, \ldots$ have to be thought of as being adjoined to the ground field. The norm D_Σ becomes a polynomial of degree nh, if η itself was a polynomial of degree h. If, e.g., instead of the element η, we consider the polynomial $z - \eta$, it follows from (5) that

$$(z-\eta)\,\omega_j = z\omega_j - \sum_k c_{jk}\omega_k,$$

(14) $$N(z-\eta) = \begin{vmatrix} z-c_{11} & -c_{12} & \cdots & -c_{1n} \\ -c_{21} & z-c_{22} & \cdots & -c_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ -c_{n1} & -c_{n2} & \cdots & z-c_{nn} \end{vmatrix}$$

The coefficient of z^{n-1} in $N(z-\eta)$ taken with the negative sign is again the trace of η:

(15) $$S(\eta) = c_{11} + c_{22} + \cdots + c_{nn}.$$

We now prove that, on the basis of the definitions (13) and (15), formulae (2) hold generally for norms and traces.

Let $(\omega_1, \ldots, \omega_n)$ be a field basis and let

(16) $$\alpha\omega_j = \sum_k a_{jk}\omega_k,$$

(17) $$\beta\omega_j = \sum_k b_{jk}\omega_k.$$

If we multiply (17) by α on the left and on the right, and if we use (16), it follows that

$$\alpha\beta\omega_j = \sum_k \sum_l b_{jk} a_{kl} \omega_l.$$

So the coefficients c_{jl} belonging to $\alpha\beta$ become:

$$c_{jl} = \sum_k b_{jk} a_{kl}.$$

By the multiplication theorem for determinants it follows that

$$N(\alpha\beta) = N(\alpha)\,N(\beta).$$

The proof of the trace relation is even easier. Addition of (16) and (17) yields

$$(\alpha+\beta)\,\omega_j = \sum_k (a_{jk} + b_{jk})\,\omega_k$$

$$S(\alpha+\beta) = \sum_j (a_{jj} + b_{jj}) = \sum_j a_{jj} + \sum_j b_{jj} = S(\alpha) + S(\beta).$$

The norm and the trace depend not only on η and the ground field Δ, but also on the extension field Σ. In order to express this, we write N_Σ or $N_{\Sigma/\Delta}$ instead of N. By (11), we have

$$N_\Omega(\eta) = N_\Sigma(\eta)^s; \quad s = (\Omega : \Sigma). \tag{18}$$

If we replace η by $z - \eta$ and compare the coefficients of the second highest power of z on the left and on the right, it follows that

$$S_\Omega(\eta) = s \cdot S_\Sigma(\eta). \tag{19}$$

We prove two more theorems, which we shall need in the following section:

1. A relation of the form

$$Nf(x_1, x_2, \ldots) = F(x_1, x_2, \ldots)$$

remains valid when the indeterminates $x_1, x_2, \ldots$ are replaced by elements of the field Δ (or by any polynomials over Δ).

2. If η is a polynomial with coefficients in Σ, the norm $N\eta$ is divisible by η.

1. follows at once from the definition of a norm (13).

2. is proved as follows: Since $N(z-\eta) = g(z)^r$ and $g(\eta) = 0$, $N(z-\eta)$ is divisible by $z-\eta$:

$$N(z-\eta) = (z-\eta)\, h(z).$$

If we put $z = 0$, the assertion follows.

EXERCISES. 1. Calculate the norm of $a + b\sqrt{d}$ in the quadratic field $\Delta(\sqrt{d})$.

2. Calculate the norm of $1 + \vartheta^2$ in the field $\Delta(\vartheta)$ if Δ is the rational number field, and if ϑ satisfies the irreducible equation $\vartheta^3 + \vartheta + 1 = 0$.

However nice the properties of the above defined norm may be, and however useful it may be for commutative fields, it is still not the same as that norm which is most frequently used for skew fields and matrix rings. For example, when we deal with a field of quaternions, the norm of $a + bj + ck + dl$ is, by our definition, equal to $(a^2 + b^2 + c^2 + d^2)^2$ whereas by the norm of a quaternion we generally mean just $a^2 + b^2 + c^2 + d^2$. Therefore, we introduce the concept of the reduced norm which is defined as follows:

Let $(\omega_1, \ldots, \omega_n)$ be a basis of the skew field Σ (or, more generally, of the hypercomplex system) under consideration. We adjoin n indeterminates $u_1, \ldots, u_n$ to the ground field Δ n, and form the "generic element" of Σ, viz.

$$\eta_u = u_1\omega_1 + \ldots + u_n\omega_n.$$

The generic element η_u is algebraic over $\Delta(u_1, \ldots, u_n)$ and, therefore, satisfies an equation of lowest degree $g_u(\eta_u) = 0$. As above, we have

$$N(z - \eta_u) = g_u(z)^r. \tag{20}$$

Since, by (14), $N(z-\eta_u)$ is a homogeneous form in $z, u_1, \ldots, u_n$, $g_u(z)$ is also such a form (with coefficients in Δ). We arrange $g_u(z)$ according to powers of x,

$$g_u(z) = z^m + b_1(u)\, z^{m-1} + \ldots + b_m(u)$$

and call $-b_1(u)$ the *reduced trace*, $(-1)^m b_m(u)$ the *reduced norm* of the generic element η_u. If

$$\eta = a_1\omega_1 + \ldots + a_n\omega_n$$

is any element of Σ, we specialize $u_k \to a_k$, and call $-b_1(a)$ the *reduced trace* and $(-1)^m b_m(a)$ the *reduced norm* of η. In symbols: $s(\eta)$ and $n(\eta)$. For $u_k = a_k$ and $z = 0$ follows from (20) that

$$N(\eta) = n(\eta)^r. \tag{21}$$

Thus the "regular norm" $N(\eta)$ is a power of the reduced norm $n(\eta)$. In case of a simple (commutative) extension field, $N(\eta)$ and $n(\eta)$ coincide. For the traces we have in analogous manner:

$$S(\eta) = r \cdot s(\eta). \tag{22}$$

Owing to the homogeneity properties of the form $g_u(z)$, the reduced trace $s(\eta_u)$ is a linear form in $u_1, \ldots, u_n$, whence the additivity of this trace

$$s(\alpha + \beta) = s(\alpha) + s(\beta)$$

follows at once. In order to prove the corresponding property of the norm, we form by means of new indeterminates $v_1, \ldots, v_n$ a second generic element

$$\eta_v = v_1 \omega_1 + \ldots v_n \omega_n.$$

Then $n(\eta_u \eta_v)$ is a form in the u and the v, and we have

$$N(\eta_u \eta_v) = N(\eta_u) N(\eta_v)$$

or, by (21),

$$n(\eta_u \eta_v)^r = n(\eta_u)^r n(\eta_v)^r;$$

hence

$$n(\eta_u \eta_v) = \gamma n(\eta_u) n(\eta_v),$$

where γ is a constant independent of u and v. If we specialize u and v so that $\eta_u = \eta_v = 1$, it follows that

$$1 = \gamma \cdot 1 \cdot 1$$

and so $\gamma = 1$. Hence we have

$$n(\eta_u) n(\eta_v) = n(\eta_u \eta_v)$$

and this property is preserved for any specialization of the u and the v.

Therefore, formulae (2) are also valid for the reduced traces and norms.

EXERCISES. 3. Prove that the reduced norm of the quaternion $a + bj + ck + dl$ is actually equal to $a^2 + b^2 + c^2 + d^2$

4. The reduced norm (in commutative fields) is equal to the regular norm if, and only if, the extension is simple.

42. THE FIELD-THEORETICAL OPERATIONS IN A FINITE NUMBER OF STEPS

In this chapter we have given various existence proofs of fields, such as the existence of the rational function field $\Delta(x_1 y_1 \ldots)$ of simple algebraic extensions $\Delta(\vartheta)$ (where ϑ satisfies a given irreducible equation), the existence of the decomposition field $\Delta(\alpha_1, \ldots, \alpha_n)$ of a polynomial $f(x)$. Now the question arises whether all these fields can actually be constructed, and whether the given existence proofs can be replaced by constructive (finite, intuitionistic) existence proofs.

We say "a field Δ is given explicitly" if its elements are uniquely represented by distinguishable symbols with which addition, subtraction, multiplication, and division can be performed in a finite number of operations.

First we show:

If the field Δ is given explicitly, then every simple transcendental extension $\Delta(t)$, as well as every simple algebraic extension $\Delta(\vartheta)$ with a given defining equation $\varphi = 0$ (which of course must be irreducible) is given explicitly.

PROOF. The elements of $\Delta(t)$ are quotients $\frac{f(t)}{g(t)}$, $g(t) \neq 0$, whose rules of operation were already set forth in Section 13. In order to force uniqueness of representation, we can demand that f and g be relatively prime, and that the leading coefficient of $g(t)$ be equal to 1. For given f and g the greatest common divisor, and thus the reduction to the relative prime form, may actually be found by means of the Euclidean algorithm (Section 18).

In the case $\Delta(\vartheta)$, the field elements may be uniquely represented in terms of $a_0 + a_1\vartheta + \cdots + a_{n-1}\vartheta^{n-1}$, and with these expressions we operate just as with polynomials, except that, at the end, the result is always reduced to the smallest remainder modulo φ (Section 32). In Section 9 it was already shown how to perform the division of two such expressions (solution of the equation $\bar{a}\,\bar{x} = \bar{b}$ in the residue class ring modulo a prime element).

In attempting to construct the decomposition field of a polynomial by the methods of Section 35 in a similar manner, we face the problem of factoring a polynomial in a given or already constructed field. There is no universal method by which, for any explicitly known field K, a decomposition into prime factors of the polynomials in $\mathsf{K}[z]$ could be performed in a finite number of steps, and there are reasons for the assumption that such a general method is impossible.[14] On the other hand, we have seen that there are such methods for factorizations into primes for special fields (fields of rationals, fields of the Gaussian integers, of the rational functions in n indeterminates with rational coefficients, of the residue classes modulo p, etc.).

1. *If, in the field* Δ, *polynomials in one indeterminate can be factored in a finite number of steps, so can polynomials in n indeterminates.*

2. *If, in the field* Δ, *a factorization of the polynomials (in one or several indeterminates) can be performed in a finite number of steps, then this is also possible for any simple transcendental extension* $\Delta(t)$ *and for any separable simple algebraic extension* $\Delta(\vartheta)$ *with a given defining equation* $\varphi(\vartheta) = 0$.

The *proof for* 1. rests on a device by Kronecker. If the polynomial $f(x_1, \ldots, x_n)$ is given, take a number m larger than the degree of the polynomial and substitute

$$x_\nu = t^{m^{\nu-1}}.$$

Here a term $a_\varrho\, x_1^{\varrho_1} x_2^{\varrho_2} \cdots x_n^{\varrho_n}$ goes into $a_\varrho\, t^{\varrho_1 + \varrho_2 m + \cdots + \varrho_n m^{n-1}}$, so different power products of the x go into different powers of t (for an integer may, in at most one way, be written in the form $\varrho_1 + \varrho_2 m + \cdots + \varrho_n m^{n-1}$, where $0 \leqq \varrho_\nu < m$). Now, if the given polynomial $f(x_1, \ldots)$ may be factored,

$$f = g_1 g_2,$$

then the factorization is also possible after the above substitution; we thus get

$$f^*(t) = g_1^*(t)\, g_2^*(t)$$

---

[14] See B. L. Van der Waerden: "Eine Bemerkung über die Unzerlegbarkeit von Polynomen". *Math. Ann.* Vol. 102 (1930), p. 738.

From the polynomials g_1^*, g_2^* we may go back to the original ones g_1, g_2, since to every term in g_ν^* only one term in g_ν can correspond. Therefore, factor $f^*(t)$ in *all possible* ways into two factors g_1^*, g_2^*, (this is possible if the decomposition into prime factors is known) and investigate each time which polynomials g_1, g_2 might belong to the factors found. Then we can check each time whether the equation

$$f = g_1 g_2$$

is correct for these g_1, g_2. If it is never correct, then f is indecomposable; if it is correct, then investigate g_1 and g_2 as before until, after at most m steps, the complete decomposition of f into prime factors is found.

The *proof of* 2. is quite simple in case of a transcendental extension $\Delta(t)$. Any polynomial $f(t, z)$ in $\Delta(t)[z]$, upon multiplication by a polynomial in t alone, can be reduced to a polynomial in t and z, and to every decomposition of $f(t, z)$ in $\Delta(t)[z]$ there corresponds, by Section 23, a decomposition in $\Delta[t, z]$, and vice versa. Thus everything has been reduced to the polynomial decomposition in $\Delta[t, z]$.

In case of an algebraic extension $\Delta(\vartheta)$, where ϑ satisfies a separable irreducible equation $\varphi(\vartheta) = 0$, the proof is not quite so simple. Let a polynomial $f(\vartheta, z)$ in $\Delta(\vartheta)[z]$ be given. To avoid ambiguities in the factorization, suppose the polynomial as well as all its divisors—if there are any—are normed in such fashion that the highest coefficient is 1. With an indeterminate u we now form the polynomial $f(\vartheta, z - u\vartheta)$, and consider its norm according to Section 41. It is again a normed polynomial

$$Nf(\vartheta, z - u\vartheta) = F(z, u).$$

In $\Delta[z, u]$ the norm can be decomposed into normed irreducible factors:

$$F(z, u) = F_1(z, u) F_2(z, u) \cdots F_r(z, u). \tag{1}$$

We now determine the greatest common divisor of each factor $F_k(z, u)$ with $f(\vartheta, z - u\vartheta)$ in $\Delta(\vartheta)[z]$. If this greatest common divisor is neither equal to $f(\vartheta, z - u\vartheta)$ nor to 1, it is a proper divisor of $f(\vartheta, z - u\vartheta)$; in this case $f(\vartheta, z - u\vartheta)$ and so $f(\vartheta, z)$ can be decomposed into two factors. We take one of these factors, say f_1, and examine it in like manner by decomposing the norm $Nf_1(\vartheta, z - u\vartheta)$. The process is completed only when each of the factors $F_k(z, u)$ is either divisible by $f(\vartheta, z - u\vartheta)$, or relatively prime to $f(\vartheta, z - u\vartheta)$. But not all factors in (1) can be relatively prime to $f(\vartheta, z - u\vartheta)$, otherwise their product $F(z, u)$ would also be relatively prime to $f(\vartheta, z - u\vartheta)$, which is impossible, since the norm of a polynomial is divisible by this same polynomial. Thus some factor of $F(z, u)$, say F_1, is divisible by $f(\vartheta, z - u\vartheta)$.

On this supposition we now assert that $f(\vartheta, z)$ is itself irreducible so that the decomposition is already completed. Suppose $f(\vartheta, z)$ is decomposable, viz.

$$f(\vartheta, z) = f_1(\vartheta, z) f_2(\vartheta, z).$$

Replacing z by $z - u\vartheta$ and forming the norm, we would get

$$F(z, u) = Nf(\vartheta, z - u\vartheta) = Nf_1(\vartheta, z - u\vartheta) Nf_2(\vartheta, z - u\vartheta). \tag{2}$$

Comparing this decomposition with (1) we see that one of the two norms on the right side in (2), say $Nf_1(\vartheta, z-u\vartheta)$ would be divisible by $F_1(z,u)$, and so by $f(\vartheta, z-u\vartheta)$:

$$Nf_1(\vartheta, z-u\vartheta) = f(\vartheta, z-u\vartheta)\, g(\vartheta, z, u).$$

Comparing on the left and on the right those terms which have the highest degree in z and in u, we obtain

$$[N(z-u\vartheta)]^m = N[(z-u\vartheta)^m] = (z-u\vartheta)^{m_1} g_1(\vartheta, z, u).$$

since f_1 and f are normed and, therefore, begin with the terms $(z-u\vartheta)^{m_1}$ and $(z-u\vartheta)^m$, respectively. Since $m > m_1$ ($f = f_1 f_2$ of higher degree than f_1), the norm $N(z-u\vartheta)$ on the left must be divisible by a power higher than the first power of $(z-u\vartheta)$:

$$N(z-u\vartheta) = (z-u\vartheta)^2 h(\vartheta, z, u).$$

Substituting $u = 1$, we obtain

$$N(z-\vartheta) = (z-\vartheta)^2 h(\vartheta, z, 1).$$

But $N(z-\vartheta) = \varphi(z)$ is the defining polynomial of the separable field $\Delta(\vartheta)$ and, therefore, has no multiple factors. Thus we have been led to a contradiction. This proves that $f(\vartheta, z)$ is irreducible under the assumptions made.

Thus it is indeed possible to find the irreducible factors of $f(\vartheta, z)$ in a finite number of steps by means of the method given (at least theoretically; for in practice the calculation and factorization of $Nf(\vartheta, z-u\vartheta)$ leads to almost insurmountable arithmetic difficulties even in the simplest cases).

The above proof is furnished in such a way that the existence of a decomposition field is not presupposed in it.

Now, if a field Δ is successively extended by the adjunction of (transcendental or separable algebraic) quantities $\vartheta_1, \vartheta_2, \ldots, \vartheta_k$, then, by the above theorems, the factorization of polynomials in the field $\Delta(\vartheta_1, \vartheta_2, \ldots, \vartheta_k)$ can be reduced step by step to that of polynomials in Δ.

Thus, the tools have been provided which are necessary to follow the existence proofs of this chapter constructively in the most important cases. We are now able to construct the decomposition field of a polynomial $f(x)$, the least normal field over a given field $\Delta(\vartheta_1, \ldots, \vartheta_k)$, the primitive element ϑ of a given field, as well as the isomorphisms of the field $\Delta(\vartheta_1, \ldots, \vartheta_k) = \Delta(\vartheta)$ in the least normal field containing it.

CHAPTER VI

CONTINUATION OF THE GROUP THEORY

CONTENTS. In Sections 43 and 44 an extension of the concept is discussed. Sections 45 to 47 contain fundamental theorems on normal divisors and "composition series," while in Sections 48 to 49 more special theorems on permutation groups will be treated. The latter will be used in the Galois theory only.

43. GROUPS WITH OPERATORS

In this section we shall extend the group concept, thus giving all subsequent investigations a greater generality which will be necessary for later applications (Chapters XV-XVII). Those readers who, for the time being, are only interested in the Galois theory may well skip this and the following section and simply think of (say finite) groups as defined previously.

Let there be given: *first*, a group $\mathfrak{G}$ (in the ordinary sense) containing elements $a, b, \ldots$; *second*, a set Ω of new objects $\eta, \Theta, \ldots$, which we shall call *operators*. Let for every Θ and every a a product Θa be defined ("the operator Θ applied to the group element a"); let this product belong to the group $\mathfrak{G}$. Furthermore, it is assumed that every single operator Θ is "distributive", i.e., that

$$\Theta(ab) = \Theta a \cdot \Theta b . \tag{1}$$

In other words: The "multiplication" by the operator Θ shall be an endormorphism of the group $\mathfrak{G}$.[1] If all these conditions are satisfied, $\mathfrak{G}$ is called a *group with operators*, and Ω the *domain of operators*.

By an *admissible subgroup* of $\mathfrak{G}$ (relative to the domain of operators Ω) we shall mean a subgroup $\mathfrak{H}$ which again admits the operators of Ω; i.e., if a belongs to $\mathfrak{H}$, every Θa shall belong to $\mathfrak{H}$. If the admissible subgroup is at the same time a normal divisor, we speak of an admissible normal divisor.

Examples: 1. Let the operators be the inner automorphisms of $\mathfrak{G}$,

$$\Theta a = cac^{-1}.$$

Admissible subgroups are the normal divisors.

[1] This implies that upon "multiplication" with Θ the identity goes into the identity, and the inverse into the inverse.

2. Let all the 1-automorphisms of $\mathfrak{G}$ be the operators. Admissible are those subgroups which, under every 1-automorphism, are transformed into themselves. These subgroups are called *characteristic subgroups.*

3. Let $\mathfrak{G}$ be a ring considered as a group under addition. Let the same ring be the domain of operators Ω; let the product Θa simply be the ring product. Then (1) is the ordinary distributive law:

$$r(a+b)=ra+rb.$$

Admissible subgroups are the *left ideals*, i.e., those subgroups which, together with any a, contain all multiples ra as well.

4. Sometimes it is advantageous to write the operators Θ on the right of the group elements, i.e., $a\Theta$ instead of Θa. In this case (1) reads:

$$(ab)\Theta=a\Theta\cdot b\Theta.$$

For example, if we regard the elements of a ring (considered as a group under addition) as right operators, where $a\Theta$ shall again be the ring product, we obtain the *right ideals* as admissible subgroups.

5. Finally, we may write part of the operators on the left, another part on the right. For example, if we take as operators for a ring the elements of the ring considered as left multipliers, as well as the same elements as right multipliers, we obtain the *two-sided* ideals as admissible subgroups.

6. By a *module* we mean any additive Abelian group. A module can equally well have a domain of operators which in this case is also called a *domain of multipliers.* We have

$$\Theta(a+b)=\Theta a+\Theta b.$$

In most cases it is assumed that the domain of multipliers is a *ring* and that

$$\text{(2)}\qquad \begin{cases}(\eta+\Theta)a=\eta a+\Theta a,\\ (\eta\Theta)a=\eta(\Theta a)\end{cases}$$

(or, with the multipliers written on the right: $a(\eta\Theta)=(a\eta)\Theta$). This implies $(\eta-\Theta)a=\eta a-\Theta a$ and $0\cdot a=0$ (the first zero is the zero element of the ring, and the second is the zero element of the module). If $\mathfrak{o}$ is the ring, we speak of $\mathfrak{o}$-*modules* or *modules with respect to the ring* $\mathfrak{o}$. If the ring has an identity ε, we very often assume that the identity is at the same time "unity operator," i.e., that $\varepsilon\cdot a=a$ for all a in $\mathfrak{G}$.

7. Every module admits as operators the ordinary integers n; for we have

$$n(a+b)=na+nb.$$

As long as we do not introduce other operators, all submodules are admissible.

8. The totality of all endormorphisms of an Abelian group (i.e., of all homomorphic mappings of the group into itself) is a domain of operators which becomes a ring if sum and product of two homomorphisms are defined by formulae (2) (in which the plus sign on the right denotes the law of combination of the group elements). This ring is called the *automorphism ring,* or better, *the endomorphism ring* of the Abelian group.

All these examples demonstrate how far-reaching the applications of groups with operators are. For further examples see the chapter on "Linear Algebra" (Volume II)

EXERCISES. 1. The intersection of two admissible subgroups is itself an admissible subgroup; the same is true for admissible normal divisors.

2. The product $\mathfrak{AB}$ of two mutually interchangeable admissible subgroups is again an admissible subgroup. For modules we have in particular: The sum $(\mathfrak{A}, \mathfrak{B})$ of two admissible submodules is itself an admissible submodule.

44. OPERATOR ISOMORPHISM AND OPERATOR HOMOMORPHISM

If $\mathfrak{G}$ and $\bar{\mathfrak{G}}$ are groups having the same domain of operators Ω, and if a mapping of $\mathfrak{G}$ upon a subset of $\bar{\mathfrak{G}}$ is given such that to every a there corresponds an $\bar{a}$, and to a product ab corresponds the product $\bar{a}\bar{b}$ and to Θa again $\Theta\bar{a}$, the mapping is called an *operator homomorphism.* If the image set is the entire group $\bar{\mathfrak{G}}$, i.e., if every element of $\bar{\mathfrak{G}}$ belongs to at least one element of $\mathfrak{G}$, we have a homomorphic mapping of $\mathfrak{G}$ upon $\bar{\mathfrak{G}}$; in the general case, however, we have a homomorphic mapping of $\mathfrak{G}$ into $\bar{\mathfrak{G}}$. If to every $\bar{a}$ there corresponds exactly one a, we have an *operator isomorphism.* We write $\mathfrak{G}\sim\bar{\mathfrak{G}}$ in case of an operator homomorphism, $\mathfrak{G}\cong\bar{\mathfrak{G}}$ in case of an operator isomorphism.

If $\mathfrak{N}$ is an admissible normal divisor of $\mathfrak{G}$, then, on applying the operator Θ, the elements ab of a coset $\bar{a}=a\mathfrak{N}$ go into $\Theta a\cdot\Theta b$, i.e., into elements of the coset $\Theta a\cdot\mathfrak{N}$. This coset $\overline{\Theta a}$ will be called the product of the operator Θ and the coset $\bar{a}$. *In this way the factor group $\mathfrak{G}/\mathfrak{N}$ becomes a group with the same domain of operators Ω, and the mapping $a\to\bar{a}$ is an operator homomorphism.*

If, on the other hand, we start with an operator homomorphism, we obtain, as in Section 10, the *law of homomorphism:*

If $\mathfrak{G}\sim\bar{\mathfrak{G}}$, then the set $\mathfrak{N}$ of the elements of $\mathfrak{G}$, to which the identity of $\bar{\mathfrak{G}}$ corresponds, is an admissible normal divisor in $\mathfrak{G}$, and the elements of $\bar{\mathfrak{G}}$ correspond to the cosets of $\mathfrak{N}$ in a one-to-one manner. Thus, the operator isomorphism

$$\mathfrak{G}/\mathfrak{N}\cong\bar{\mathfrak{G}}.$$

holds.

We already know from Section 10 that $\mathfrak{N}$ is a normal divisor. It is obvious that $\mathfrak{N}$ is admissible; for if a is mapped upon the identity $\bar{e}$, Θa is mapped upon $\Theta\bar{e}=\bar{e}$, i.e. if a belongs to $\mathfrak{N}$, so does Θa. We already know that the correspondence of the cosets to the elements of $\bar{\mathfrak{G}}$ is one-to-one. At the same time it is an operator isomorphism, since the given correspondence $\mathfrak{G}\to\bar{\mathfrak{G}}$ is an operator homomorphism.

For additive groups with a domain of operators $\mathfrak{o}$ ($\mathfrak{o}$-modules, in particular, ideals in $\mathfrak{o}$) the operator homomorphism is also called a *module homomorphism.*

We note that in such a homomorphism Θa goes into $\Theta \bar{a}$ so that Θ remains untransformed; this is the difference between a module homomorphism and a ring homomorphism, in which ab is transformed into $\bar{a}\bar{b}$. For example, two left ideals in a ring $\mathfrak{o}$ may be regarded as $\mathfrak{o}$-modules; then an operator homomorphism associates an $\bar{a}$ with every a, and the product $r\bar{a}$ with the product ra (where r is in $\mathfrak{o}$). They may also be thought of as ring; a ring homomorphism associates with the product ra (r in the ideal) not $r\bar{a}$, but $\bar{r}\bar{a}$.

Whenever we speak hereafter of "groups," we shall include groups with operators as well. By "subgroups" and "normal divisors" we shall tacitly understand admissible subgroups and normal divisors, and by "isomorphism" and "homomorphism," operator isomorphism and operator homomorphisms. In the following two sections we shall exclusively use the law of homomorphism and the facts that the intersection of two (admissible) subgroups is itself an admissible subgroup, and that the product of two mutually interchangeable subgroups (in particular the product of a normal divisor by a subgroup) is itself an admissible subgroup.

EXERCISES. 1. The ideals (1) and (2) in the ring of integers are isomorphic modules, but not isomorphic rings.

2. In the ring of the number pairs (a_1, a_2) (Section 11, Ex. 1) the ideals generated by (1, 0) and (0, 1) are isomorphic rings, but not isomorphic modules.

45. THE TWO LAWS OF ISOMORPHISM

In the homomorphism $\mathfrak{G} \sim \overline{\mathfrak{G}} = \mathfrak{G}/\mathfrak{N}$ every subgroup $\mathfrak{H}$ of $\mathfrak{G}$ is homomorphically mapped upon a subgroup $\overline{\mathfrak{H}}$ of $\overline{\mathfrak{G}}$. Now, if we go back from $\overline{\mathfrak{H}}$ and wish to find in $\mathfrak{G}$ the totality $\mathfrak{K}$ of those elements whose images (or cosets) belong to $\overline{\mathfrak{H}}$, it may happen that $\mathfrak{K}$ includes more elements than that of $\mathfrak{H}$; for with every a in $\mathfrak{H}$, $\mathfrak{K}$ contains all elements of the coset $a\mathfrak{N}$. If we denote by $\mathfrak{H}\mathfrak{N}$ that group which consists of all products ab formed by an a in $\mathfrak{H}$ and a b in $\mathfrak{N}$ (cf. Ex. 2, Section 43), it follows that $\mathfrak{K} = \mathfrak{H}\mathfrak{N}$, and also $\overline{\mathfrak{H}} = \mathfrak{H}\mathfrak{N}/\mathfrak{N}$. On the other hand, $\mathfrak{H}$ is homomorphically mapped upon $\overline{\mathfrak{H}}$; hence $\overline{\mathfrak{H}}$ is isomorphic with the factor group of $\mathfrak{H}$ with respect to a normal divisor of $\mathfrak{H}$ consisting of those elements of $\mathfrak{H}$ to which the identity corresponds, i.e., of those elements of $\mathfrak{H}$, which also belong to $\mathfrak{N}$. From this we have the *first law of isomorphism:*

If $\mathfrak{N}$ is a normal divisor in $\mathfrak{G}$, and $\mathfrak{H}$ a subgroup of $\mathfrak{G}$, the intersection $\mathfrak{H} \cap \mathfrak{N}$ is a normal divisor in $\mathfrak{H}$, and we have [2]

$$\mathfrak{H}\mathfrak{N}/\mathfrak{N} \cong \mathfrak{H}/(\mathfrak{H} \cap \mathfrak{N}).$$

The totality of elements mapped into $\overline{\mathfrak{H}}$ will be exactly equal to $\mathfrak{H}$ if, for

[2] For modules, of course, we have to write $(\mathfrak{H}, \mathfrak{N})$ instead of $\mathfrak{H}\mathfrak{N}$.

every a in $\mathfrak{H}$, the entire coset $a\mathfrak{N}$ is contained in $\mathfrak{H}$, i.e., if

$$\mathfrak{H} \geq \mathfrak{N}.$$

There exists a one-to-one correspondence between these groups $\mathfrak{H} \geq \mathfrak{N}$ and certain groups $\overline{\mathfrak{H}} = \mathfrak{H}/\mathfrak{N}$ in $\overline{\mathfrak{G}}$. Furthermore, *every* subgroup $\overline{\mathfrak{H}}$ of $\overline{\mathfrak{G}}$ defines a subgroup $\mathfrak{H} \geq \mathfrak{N}$ consisting of all elements of all cosets of $\mathfrak{N}$ which occur in $\overline{\mathfrak{H}}$. Finally, the right and left cosets of $\overline{\mathfrak{H}}$ in $\overline{\mathfrak{G}}$ correspond to the right and left cosets of $\mathfrak{H}$ in $\mathfrak{G}$, respectively. Hence, if $\overline{\mathfrak{H}}$ is a normal divisor in $\overline{\mathfrak{G}}$, $\mathfrak{H}$ is a normal divisor in $\mathfrak{G}$, and vice versa. A part of these results will be obtained by a different method in the proof of the *second law of isomorphism*:

If $\overline{\mathfrak{G}} = \mathfrak{G}/\mathfrak{N}$, and $\overline{\mathfrak{H}}$ is a normal divisor in $\mathfrak{G}$, then the corresponding subgroup $\mathfrak{H}$ is a normal divisor in $\mathfrak{G}$, and we have

(1) $$\mathfrak{G}/\mathfrak{H} \cong \overline{\mathfrak{G}}/\overline{\mathfrak{H}}.$$

PROOF. We have $\mathfrak{G} \sim \overline{\mathfrak{G}}$ and $\overline{\mathfrak{G}} \sim \overline{\mathfrak{G}}/\overline{\mathfrak{H}}$, and therefore $\mathfrak{G} \sim \overline{\mathfrak{G}}/\overline{\mathfrak{H}}$. Hence $\overline{\mathfrak{G}}/\overline{\mathfrak{H}}$ is isomorphic with a factor group of $\mathfrak{G}$ with respect to a normal divisor. This normal divisor consists of those elements of $\mathfrak{G}$ to which, in the homomorphism $\mathfrak{G} \sim \overline{\mathfrak{G}}/\overline{\mathfrak{H}}$, corresponds the identity, i.e., to which corresponds in the first homomorphism $\mathfrak{G} \sim \overline{\mathfrak{G}}$ an element of $\overline{\mathfrak{H}}$. This normal divisor is $\mathfrak{H}$. Q.E.D.

The 1-isomorphism may also be written thus: $\mathfrak{G}/\mathfrak{H} \cong (\mathfrak{G}/\mathfrak{N})/(\mathfrak{H}/\mathfrak{N})$.

EXERCISES. 1. Show by means of the first law of isomorphism that the factor group of the symmetric group $\mathfrak{S}_4$ with respect to the four-group $\mathfrak{B}_4$ (Section 9, Ex. 4) is isomorphic with the symmetric group $\mathfrak{S}_3$.

2. Show in like manner that in every permutation group which does not consist of only even permutation the even permutations form a normal divisor of index 2.

3. Show in like manner that the factor group of the Euclidean group of motions with respect to the normal divisor of the translations is isomorphic with the group of rotations about a point.

46. NORMAL SERIES AND COMPOSITION SERIES

A group $\mathfrak{G}$ is called *simple* if it possesses no normal divisor, except itself and the identity group.

Examples: Groups of prime order are simple, since the order of a subgroup would have to be a divisor of the order of the group. It will be shown later that the alternating group $\mathfrak{A}_n$ for $n > 4$ is simple (Section 48). Moreover, any module of rank one over a field is simple if the field is regarded as a domain of multipliers.

A finite sequence of subgroups of a group $\mathfrak{G}$,

(1) $$\{\mathfrak{G} = \mathfrak{G}_0 \geq \mathfrak{G}_1 \geq \cdots \geq \mathfrak{G}_l = \mathfrak{E}\}$$

is called a *normal series* if, for $\nu = 1, \ldots, l$ every $\mathfrak{G}_\nu$ is a normal divisor in $\mathfrak{G}_{\nu-1}$. The number l is called the *length* of the normal series; the factor groups $\mathfrak{G}_{\nu-1}/\mathfrak{G}_\nu$ are called the *factors* of the normal series. It should be noted that the length is not

the number of terms of the sequence (1), but the number of the factors $\mathfrak{G}_{\nu-1}/\mathfrak{G}_\nu$, which is one less.

A second normal series

$$\{\mathfrak{G} \geq \mathfrak{H}_1 \geq \cdots \geq \mathfrak{H}_m = \mathfrak{E}\} \tag{2}$$

is called a *refinement* of the first one if all $\mathfrak{G}_i$ in (1) also occur in (2). For example, for the group $\mathfrak{S}_4$ the series

$$\{\mathfrak{S}_4 \supset \mathfrak{A}_4 \supset \mathfrak{B}_4 \supset \mathfrak{E}\}.$$

(cf. Section 9, Ex. 4) is a refinement of

$$\{\mathfrak{S}_4 \supset \mathfrak{B}_4 \supset \mathfrak{E}\}.$$

In a normal series a term may be repeated any number of times: $\mathfrak{G}_i = \mathfrak{G}_{i+1} = \cdots = \mathfrak{G}_k$. If this does *not* happen, we speak of a series *without repetitions*. A series without repetitions which cannot be refined any more without repetitions is called a *composition series*. For example, in the symmetric group $\mathfrak{S}_3$ the series

$$\{\mathfrak{S}_3 \supset \mathfrak{A}_3 \supset \mathfrak{E}\}$$

is a composition series, likewise in $\mathfrak{S}_4$ the series

$$\{\mathfrak{S}_4 \supset \mathfrak{A}_4 \supset \mathfrak{B}_4 \supset \{1, (1\,2)\,(3\,4)\} \supset \mathfrak{E}\}.$$

In either case the impossibility of further refinements is seen from the fact that the index of each of the successive normal divisors in the preceding one is prime. However, there are groups in which every normal series can be further refined; such groups, therefore, do not possess a composition series. An example of such a case is any infinite cyclic group; for if in such a group a normal series without repetitions

$$\{\mathfrak{G} \supset \mathfrak{G}_1 \supset \cdots \supset \mathfrak{G}_{l-1} \supset \mathfrak{E}\}$$

is given, and if $\mathfrak{G}_{l-1}$ has, e.g., the index m so that $\mathfrak{G}_{l-1} = \{a^m\}$, there always exists a subgroup $\{a^{2m}\}$ of index $2m$ between $\mathfrak{G}_{l-1}$ and $\mathfrak{E}$.

A normal series is a composition series if, and only if, between every two succeeding terms $\mathfrak{G}_{\nu-1}$ and $\mathfrak{G}_\nu$ no normal divisor of $\mathfrak{G}_{\nu-1}$ distinct from these terms can be interpolated, or, what is the same thing according to Section 45, if $\mathfrak{G}_{\nu-1}/\mathfrak{G}_\nu$ is simple. The simple factors $\mathfrak{G}_{\nu-1}/\mathfrak{G}_\nu$ of a composition series are called *composition factors*. In our examples the composition factors are cyclic groups of orders 2, 3, and 2, 3, 2, 2, respectively.

Two normal series are called *isomorphic* if all factors $\mathfrak{G}_{\nu-1}/\mathfrak{G}_\nu$ of one series are 1-isomorphic with the factors of the second series in any sequential order. For example, in a cyclic group $\{a\}$ of order 6 the two series

$$\{\{a\},\ \{a^2\},\ \mathfrak{E}\},$$
$$\{\{a\},\ \{a^3\},\ \mathfrak{E}\}$$

are isomorphic; for the factors of the first series are cyclic and of orders 2, 3, and those of the second series are cyclic and of orders 3, 2.—For the sake of convenience, we shall hereafter use the symbol $\cong$ also for the isomorphism of normal series.

If a chain of normal divisors

$$\{\mathfrak{G} \geq \mathfrak{G}_1 \geq \cdots\}$$

terminates with any normal divisor $\mathfrak{A}$ of $\mathfrak{G}$, which need not be equal to $\mathfrak{E}$, we speak of a *normal series from* $\mathfrak{G}$ to $\mathfrak{A}$; to such a series there corresponds a normal series

$$\{\mathfrak{G}/\mathfrak{A} \geq \mathfrak{G}_1/\mathfrak{A} \geq \cdots \geq \mathfrak{A}/\mathfrak{A} = \mathfrak{E}\}$$

of the factor group $\mathfrak{G}/\mathfrak{A}$, and vice versa. By the second law of isomorphism, the factors of the second series are isomorphic with those of the first.

If two normal series

$$\{\mathfrak{G} \geq \mathfrak{G}_1 \geq \cdots \geq \mathfrak{G}_r = \mathfrak{E}\}$$

and

$$\{\mathfrak{G} \geq \mathfrak{H}_1 \geq \cdots \geq \mathfrak{H}_r = \mathfrak{E}\}$$

are isomorphic, then, for every refinement of the first series, an isomorphic refinement of the second series can be found. For every factor $\mathfrak{G}_{\nu-1}/\mathfrak{G}_\nu$ is isomorphic with a definite factor $\mathfrak{H}_{\mu-1}/\mathfrak{H}_\mu$; thus to every normal series for $\mathfrak{G}_{\nu-1}/\mathfrak{G}_\nu$ there corresponds an isomorphic normal series for $\mathfrak{H}_{\mu-1}/\mathfrak{H}_\mu$ and, therefore, to every normal series from $\mathfrak{G}_{\nu-1}$ to $\mathfrak{G}_\nu$ there corresponds an isomorphic series from $\mathfrak{H}_{\mu-1}$ to $\mathfrak{H}_\mu$.

We can now prove the following *Fundamental Theorem* on *Normal Series*, due to O. Schreier: *Two arbitrary normal series of an arbitrary group* $\mathfrak{G}$,

$$\{\mathfrak{G} \geq \mathfrak{G}_1 \geq \mathfrak{G}_2 \geq \cdots \geq \mathfrak{G}_r = \mathfrak{E}\},$$
$$\{\mathfrak{G} \geq \mathfrak{H}_1 \geq \mathfrak{H}_2 \geq \cdots \geq \mathfrak{H}_s = \mathfrak{E}\}$$

possess isomorphic refinements

$$\{\mathfrak{G} \geq \cdots \geq \mathfrak{G}_1 \geq \cdots \geq \mathfrak{G}_2 \geq \cdots \geq \mathfrak{E}\}$$
$$\cong \{\mathfrak{G} \geq \cdots \geq \mathfrak{H}_1 \geq \cdots \mathfrak{H}_2 \geq \cdots \geq \mathfrak{E}\}.$$

PROOF. For $r = 1$ or $s = 1$ the proof is clear; for in this case one of the series is $\{\mathfrak{G} \geq \mathfrak{E}\}$, and the other is automatically a refinement of it.

Let us first prove the theorem for $s = 2$ by the method of complete induction on r, and next for arbitrary s by complete induction on s.

For $s = 2$ the second series reads:

$$\{\mathfrak{G} \geq \mathfrak{H} \geq \mathfrak{E}\}.$$

We put $\mathfrak{D} = \mathfrak{G}_1 \cap \mathfrak{H}$ and $\mathfrak{P} = \mathfrak{G}_1\,\mathfrak{H}$; then $\mathfrak{P}$ and $\mathfrak{D}$ are normal divisors in $\mathfrak{G}$. Of course it is possible that $\mathfrak{P} = \mathfrak{G}$ or $\mathfrak{D} = \mathfrak{E}$. Under the induction hypothesis the series of lengths $r - 1$ and 2, namely

$$\{\mathfrak{G}_1 \geq \mathfrak{G}_2 \geq \cdots \geq \mathfrak{G}_r = \mathfrak{E}\} \text{ and } \{\mathfrak{G}_1 \geq \mathfrak{D} \geq \mathfrak{E}\}$$

possess isomorphic refinements

$$\text{(3)} \quad \left\{ \begin{array}{l} \{\mathfrak{G}_1 \geq \cdots \geq \mathfrak{G}_2 \geq \cdots \geq \mathfrak{E}\} \\ \qquad\qquad \cong \{\mathfrak{G}_1 \geq \cdots \geq \mathfrak{D} \geq \cdots \geq \mathfrak{E}\}. \end{array} \right.$$

By the first law of isomorphism we have, moreover,

$$\mathfrak{P}/\mathfrak{H} \cong \mathfrak{G}_1/\mathfrak{D} \text{ and } \mathfrak{P}/\mathfrak{G}_1 \cong \mathfrak{H}/\mathfrak{D}$$

so that

$$\{\mathfrak{P} \geq \mathfrak{G}_1 \geq \mathfrak{D} \geq \mathfrak{E}\} \cong \{\mathfrak{P} \geq \mathfrak{H} \geq \mathfrak{D} \geq \mathfrak{E}\}. \tag{4}$$

The right side of (3) yields a refinement of the left side of (4), for which an isomorphic refinement of the right side can be found:

$$\left| \begin{array}{l} \{\mathfrak{P} \geq \mathfrak{G}_1 \geq \cdots \geq \mathfrak{D} \geq \cdots \geq \mathfrak{E}\} \\ \qquad \cong \{\mathfrak{P} \geq \cdots \geq \mathfrak{H} \geq \mathfrak{D} \geq \cdots \geq \mathfrak{G}\}. \end{array} \right. \tag{5}$$

From (3) and (5) we have

$$\left| \begin{array}{l} \{\mathfrak{G} \geq \mathfrak{P} \geq \mathfrak{G}_1 \geq \cdots \geq \mathfrak{G}_2 \geq \cdots \geq \mathfrak{E}\} \\ \qquad \cong \{\mathfrak{G} \geq \mathfrak{P} \geq \cdots \geq \mathfrak{H} \geq \mathfrak{D} \geq \cdots \geq \mathfrak{E}\}, \end{array} \right.$$

which proves the theorem for the case $s = 2$.

For arbitrary s we may, by what has just been proved, refine the first series $\{\mathfrak{G} \geq \mathfrak{G}_1 \geq \cdots\}$ in such a fashion that it becomes isomorphic with a refinement of $\{\mathfrak{G} \geq \mathfrak{H}_1 \geq \mathfrak{E}\}$:

$$\left| \begin{array}{l} \{\mathfrak{G} \geq \cdots \geq \mathfrak{G}_1 \geq \cdots \geq \mathfrak{G}_2 \geq \cdots \geq \mathfrak{E}\} \\ \qquad \cong \{\mathfrak{G} \geq \cdots \geq \mathfrak{H}_1 \geq \cdots \geq \mathfrak{E}\}. \end{array} \right. \tag{6}$$

By the induction hypothesis, the partial series on the right $\{\mathfrak{H}_1 \geq \cdots \geq \mathfrak{E}\}$ and the series $\{\mathfrak{H}_1 \geq \mathfrak{H}_2 \geq \cdots \geq \mathfrak{H}_s = \mathfrak{E}\}$ have isomorphic refinements:

$$\{\mathfrak{H}_1 \geq \cdots \geq \mathfrak{E}\} \cong \{\mathfrak{H}_1 \geq \cdots \geq \mathfrak{H}_2 \geq \cdots \geq \mathfrak{E}\}. \tag{7}$$

The left side of (7) yields a refinement of the right side of (6) for which an isomorphic refinement of the left side of (6) can be found. Thus we haye

$$\begin{array}{ll} & \{\mathfrak{G} \geq \cdots \geq \mathfrak{G}_1 \geq \cdots \geq \mathfrak{G}_2 \geq \cdots \geq \mathfrak{E}\} \\ & \cong \{\mathfrak{G} \geq \cdots \geq \mathfrak{H}_1 \geq \cdots \geq \mathfrak{E}\} \\ [\text{by (7)}] & \cong \{\mathfrak{G} \geq \cdots \geq \mathfrak{H}_1 \geq \cdots \geq \mathfrak{H}_2 \geq \cdots \geq \mathfrak{E}\}. \end{array}$$

This completes the proof of the theorem.

If, in two isomorphic series, we strike out all repetitions, the series remain isomorphic. Thus the refinements referred to in the Fundamental Theorem may always be assumed to be without repetitions.

From the Fundamental Theorem on Normal Series we immediately infer the following two theorems for groups that possess a composition series.

1. Jordan-Hölder Theorem: *Any two composition series of one and the same group are isomorphic.*

For these series cannot be refined without repetitions.

2. *If* $\mathfrak{G}$ *possesses a composition series, every normal series of* $\mathfrak{G}$ *can be re-*

fined to a composition series; in particular, every normal divisor can be included in a composition series.[3]

A group is called *soluble* (or solvable) if it possesses a normal series in which all factors are Abelian. (Examples: the groups $\mathfrak{S}_3$ and $\mathfrak{S}_4$; see above.)

It follows from the Fundamental Theorem that in a soluble group every normal series can be refined to one with Abelian factors. If, in particular, the group has a composition series, then all composition factors are simple Abelian groups. In case of ordinary finite groups without operators, this means that they are cyclic groups of prime order. (Cf. Ex. 3 below.)

EXERCISES. 1. Every finite group possesses a composition series.

2. Form all possible composition series of a cyclic group of order 20.

3. An Abelian group (without operators) is simple only if it is cyclic of prime order.

4. A group of order p^n is simple only if $n = 1$ (cf. Section 9, Ex. 9).

5. Every group of order p^n is soluble. [Form a composition series and apply Ex. 4.]

47. DIRECT PRODUCTS

The group $\mathfrak{G}$ is called a *direct product* of the subgroups $\mathfrak{A}$ and $\mathfrak{B}$ if the following conditions are satisfied:

I. 1. $\mathfrak{A}$ and $\mathfrak{B}$ are normal divisors in $\mathfrak{G}$,

2. $\mathfrak{G} = \mathfrak{A}\mathfrak{B}$,

3. $\mathfrak{A} \cap \mathfrak{B} = \mathfrak{E}$.

Equivalent conditions to the above are:

II. 1. Every element of $\mathfrak{G}$ is expressible as a product

$$g = ab, \quad a \in \mathfrak{A}, \quad b \in \mathfrak{B} \tag{1}$$

2. the factors a and b are uniquely determined by g,

3. every element of $\mathfrak{A}$ commutes with every element of $\mathfrak{B}$.

I. *implies* II. Obviously, I.2 implies II.1 II.2 follows thus: If $g = a_1b_1 = a_2b_2$, then $a_2^{-1}a_1 = b_2b_1^{-1}$; this element $a_2^{-1}a_1$ must belong both to $\mathfrak{A}$ and to $\mathfrak{B}$; hence, by I.3, it must be equal to the identity; this implies

$$a_1 = a_2, \quad b_1 = b_2,$$

hence the uniqueness. II.3 follows from the fact that $aba^{-1}b^{-1}$ belongs both to $\mathfrak{A}$ and $\mathfrak{B}$ because of I.1, and therefore, by I.3, is the identity.

---

[3] Another proof of these two theorems was given by E. Noether: "Abstrakter Aufbau der Idealtheorie in algebraischen Zahl-und Funktionenkörpern." *Math. Ann.* Bd. 96 (1926), p. 57, Paragraph 10.

II. *implies* I. The normal divisor property of $\mathfrak{A}$ is proved thus:

$$g\,\mathfrak{A}\,g^{-1} = a\,b\,\mathfrak{A}\,b^{-1}a^{-1} = a\,\mathfrak{A}\,a^{-1} = \mathfrak{A} \qquad [\text{because of II.3}].$$

I.2 follows from II.1. Finally, I.3 is shown as follows: If c is an element of $\mathfrak{A} \cap \mathfrak{B}$, then c can be expressed as a product of an element of $\mathfrak{A}$ by an element of $\mathfrak{B}$ in two ways:

$$c = c \cdot 1 = 1 \cdot c.$$

Because of the uniqueness [II.3] we must have $c = 1$. This proves I.3.

If $\mathfrak{A}\mathfrak{B}$ is a direct product, it is also denoted by $\mathfrak{A} \times \mathfrak{B}$. In additive groups (modules) we write $(\mathfrak{A}, \mathfrak{B})$ for the sum, and $\mathfrak{A} + \mathfrak{B}$ for the direct sum.

If the structure of $\mathfrak{A}$ and $\mathfrak{B}$ is known, so is the structure of $\mathfrak{G}$; for we multiply two elements $g_1 = a_1 b_1$ and $g_2 = a_2 b_2$ by multiplying their factors:

$$g_1 g_2 = a_1 a_2 \cdot b_1 b_2.$$

The group $\mathfrak{G}$ is called the *direct product of several subgroups* $\mathfrak{G} = \mathfrak{A}_1 \times \mathfrak{A}_2 \times \cdots \times \mathfrak{A}_n$ if the following conditions are satisfied:

I'. 1. All $\mathfrak{A}_\nu$ are normal divisors in $\mathfrak{G}$,

2. $\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_n = \mathfrak{G}$,

3. $(\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{\nu-1}) \cap \mathfrak{A}_\nu = \mathfrak{E}$ $\qquad (\nu = 2, 3, \ldots, n)$.

If these conditions are fulfilled, then the groups $\mathfrak{A}_1, \ldots, \mathfrak{A}_{n-1}$ are also normal divisors in their product $\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{n-1}$ so that, by the same definition, this product is direct; furthermore, $\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{n-1}$ being a product of normal divisors, is itself a normal divisor in $\mathfrak{G}$, and we have $(\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{n-1}) \cap \mathfrak{A}_n = \mathfrak{E}$ so that

$$\mathfrak{G} = (\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{n-1}) \times \mathfrak{A}_n = \mathfrak{B}_n \times \mathfrak{A}_n \tag{2}$$

with

$$\mathfrak{B}_n = \mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{n-1} = \mathfrak{A}_1 \times \mathfrak{A}_2 \ldots \times \mathfrak{A}_{n-1}.$$

By means of (2) we may also define the direct product of n factors recursively. If we apply to $\mathfrak{G} = \mathfrak{B}_n \times \mathfrak{A}_n$ definition II., which is equivalent to I., we infer readily by the method of complete induction on n:

II'. *Every element g of $\mathfrak{G}$ is uniquely expressible as a product*

$$g = a_1 a_2 \ldots a_n \qquad (a_\nu \in \mathfrak{A}_\nu)$$

and every element of $\mathfrak{A}_\mu$ commutes with every element of $\mathfrak{A}_\nu$ $(\mu \neq \nu)$.

Conversely, from II'. follows I'; for if we put

$$\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{\nu-1} \mathfrak{A}_{\nu+1} \ldots \mathfrak{A}_n = \mathfrak{B}_\nu,$$

it follows from II'., for every ν, that

$$\mathfrak{G} = \mathfrak{A}_\nu \times \mathfrak{B}_\nu. \tag{3}$$

So every $\mathfrak{A}_\nu$ is a normal divisor in $\mathfrak{G}$, and

$$\mathfrak{A}_\nu \cap \mathfrak{B}_\nu = \mathfrak{E}.$$

The latter assertion implies even a little more than condition I'.3.

By the first law of isomorphism it follows from (3) that

$$\mathfrak{G}/\mathfrak{A}_\nu \cong \mathfrak{B}_\nu; \qquad \mathfrak{G}/\mathfrak{B}_\nu \cong \mathfrak{A}_\nu.$$

The groups

$$(4)\qquad \begin{cases} \mathfrak{G} &= \mathfrak{A}_1 \times \mathfrak{A}_2 \times \cdots \times \mathfrak{A}_n, \\ \mathfrak{G}_1 &= \mathfrak{A}_1 \times \mathfrak{A}_2 \times \cdots \times \mathfrak{A}_{n-1}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \mathfrak{G}_{n-1} &= \mathfrak{A}_1, \\ \mathfrak{G}_n &= \mathfrak{E} \end{cases}$$

form a normal series of $\mathfrak{G}$ with the factors $\mathfrak{G}_{\nu-1}/\mathfrak{G}_\nu \cong \mathfrak{A}_{n-\nu+1}$. If the groups $\mathfrak{A}_\nu$ possess composition series, then $\mathfrak{G}$ possesses a composition series [refinement of the above normal series (2)] whose length is the sum of the lengths of the individual factors.

The following theorem can be proved very easily:

If $\mathfrak{G} = \mathfrak{A} \times \mathfrak{B}$, $\mathfrak{G}'$ *a subgroup of* $\mathfrak{G}$, *and* $\mathfrak{G}' \geq \mathfrak{A}$, *then* $\mathfrak{G}' = \mathfrak{A} \times \mathfrak{B}'$, *where* $\mathfrak{B}'$ *represents the intersection of* $\mathfrak{G}'$ *and* $\mathfrak{B}$.

PROOF. For every element of $\mathfrak{G}$ a representation $g = a \cdot b$ is valid. This representation is, in particular, valid for the elements of $\mathfrak{G}'$. The factors b which occur in it are contained both in $\mathfrak{B}$ and in $\mathfrak{G}'$ (since both g and a lie in $\mathfrak{G}'$); therefore, the b belong to the intersection $\mathfrak{G}' \cap \mathfrak{B}$. On the other hand, for a and b we may choose any elements of $\mathfrak{A}$ or $\mathfrak{G}' \cap \mathfrak{B}$, respectively, and we always obtain a product $g = a \cdot b$ in $\mathfrak{G}'$. Hence $\mathfrak{G}' = \mathfrak{A} \times (\mathfrak{G}' \cap \mathfrak{B})$. Q.E.D.

EXERCISES. 1. A cyclic group $\{a\}$ of order $n = r \cdot s$ with $(r, s) = 1$ is the direct product of its subgroups $\{a^r\} \cdot \{a^s\}$ of orders s and r.

2. A finite cyclic group is the direct product of its subgroups of the highest possible prime power orders.

A group $\mathfrak{G}$ is called *completely reducible* if it is a direct product of simple groups. In this case the normal series (4) is already a composition series. By the Jordan-Hölder theorem, the composition factors $\mathfrak{G}_{\nu-1}/\mathfrak{G}_\nu \cong \mathfrak{A}_{n-\nu+1}$ are uniquely determined, except for isomorphism and sequential order.

THEOREM. *In a completely reducible group* $\mathfrak{G}$ *every normal divisor is a direct factor, i.e., for every normal divisor* $\mathfrak{H}$ *there exists a factorization* $\mathfrak{G} = \mathfrak{H} \times \mathfrak{B}$.

PROOF. From $\mathfrak{G} = \mathfrak{A}_1 \times \mathfrak{A}_2 \times \cdots \times \mathfrak{A}_n$ follows

$$(5)\qquad \mathfrak{G} = \mathfrak{H} \cdot \mathfrak{G} = \mathfrak{H} \cdot \mathfrak{A}_1 \cdot \mathfrak{A}_2 \cdots \mathfrak{A}_n.$$

Now, with each of the factors $\mathfrak{A}_1, \ldots, \mathfrak{A}_n$ we can perform an operation which consists either in striking out the factor, or in replacing the preceding symbol by the symbol $\times$ for the direct product; for the intersection of the $\mathfrak{A}_k$ under consideration with the preceding product $\Pi = \mathfrak{H} \cdot \mathfrak{A}_1 \cdots \mathfrak{A}_{k-1}$ is a normal divisor in $\mathfrak{A}_k$, so it is equal either to $\mathfrak{A}_k$ or to $\mathfrak{E}$. In the first case, $\Pi \cap \mathfrak{A}_k = \mathfrak{A}_k$, we have $\mathfrak{A}_k \subset \Pi$ so that the factor $\mathfrak{A}_k$ in the product $\Pi \mathfrak{A}_k$ is superfluous. In the second case the product $\Pi \cdot \mathfrak{A}_k$ is direct: $\Pi \cdot \mathfrak{A}_k = \Pi \times \mathfrak{A}_k$.

By what has just been proved, the product (5) assumes, after removal of all superfluous $\mathfrak{A}$, the form of a direct product:

$$\mathfrak{G} = \mathfrak{H} \times \mathfrak{A}_{i_1} \times \mathfrak{A}_{i_2} \times \cdots \times \mathfrak{A}_{i_r}.$$

This proves the proposition.

48. SIMPLICITY OF THE ALTERNATING GROUP

In Section 46 we saw that the symmetric groups $\mathfrak{S}_3$, $\mathfrak{S}_4$ are soluble. All other symmetric groups $\mathfrak{S}_n (n > 4)$, on the other hand, are not soluble. Though they always have a normal divisor of indox 2, namely the alternating group $\mathfrak{A}_n$, the composition series goes from $\mathfrak{A}_n$ to $\mathfrak{E}$ directly, according to the following

THEOREM: *The alternating group $\mathfrak{A}_n (n > 4)$ is simple.*

We need a

LEMMA. *If a normal divisor $\mathfrak{N}$ of the group $\mathfrak{A}_n (n > 2)$ contains a cyclic permutation of three digits, then $\mathfrak{N} = \mathfrak{A}_n$.*

PROOF OF THE LEMMA. Let $\mathfrak{N}$ contain the cycle (1 2 3). Then $\mathfrak{N}$ must also contain its square (2 1 3), as well as all transformed symbols

$$\sigma \cdot (2\ 1\ 3) \cdot \sigma^{-1} \qquad (\sigma \in \mathfrak{A}_n)$$

If we choose $\sigma = (1\ 2)\ (3\ k)$, where $k > 3$, we have

$$\sigma \cdot (2\ 1\ 3) \cdot \sigma^{-1} = (1\ 2\ k);$$

thus $\mathfrak{N}$ contains all cycles of the form (1 2 k). But these generate the group $\mathfrak{A}_n$ (Section 7, Ex. 4); hence we must have $\mathfrak{N} = \mathfrak{A}_n$.

PROOF OF THE THEOREM. Let $\mathfrak{N}$, distinct from $\mathfrak{E}$, be a normal divisor in $\mathfrak{A}_n$. We want to show that $\mathfrak{N} = \mathfrak{A}_n$.

We choose a permutation τ in $\mathfrak{N}$, which, without being equal to 1, leaves fixed as many digits as possible. We shall show that τ displaces exactly 3 symbols and leaves fixed all others.

Let us suppose that τ displaces more than 3 symbols; then, in the cycle representation of τ at least 4 symbols actually occur. τ either contains a cycle of at least 3 symbols, or τ consists of cycles of 2 symbols only. In the first case we put

$$\tau = (1\ 2\ 3 \ldots) \ldots .$$

In this case τ replaces at least the symbols 1, 2, 3, 4, 5; for the odd permutation (1 2 3 4) does not occur in the alternating group. In the second case we put

$$\tau = (1\ 2)(3\ 4) \ldots .$$

We now transform τ with $\sigma = (3\ 4\ 5)$ and find in the first case

$$\tau_1 = \sigma\tau\sigma^{-1} = (1\ 2\ 4 \ldots) \ldots,$$

and in the second case

$$\tau_1 = \sigma\tau\sigma^{-1} = (1\ 2)\ (4\ 5) \ldots,$$

i.e., $\tau_1 \neq \tau$ in both cases so that $\tau^{-1}\tau_1 \neq 1$. But the permutation $\tau^{-1}\tau_1$ leaves invariant more symbols than τ itself, contrary to the definition of τ. Hence τ

cannot displace more than 3 symbols. But then τ is a cyclic permutation of 3 symbols, and by our lemma we have $\mathfrak{R}=\mathfrak{A}_n$. This proves everything.

EXERCISE. Prove that, for $n\neq 4$, the alternating group $\mathfrak{A}_n$ is the only normal divisor of the symmetric group $\mathfrak{S}_n$, except the latter itself and $\mathfrak{E}$.

49. TRANSITIVITY AND PRIMITIVITY

A group of permutations of a set $\mathfrak{M}$ is called *transitive over* $\mathfrak{M}$ if there exists an element a in $\mathfrak{M}$ which the permutations of the group carry into all elements x of $\mathfrak{M}$, which means that for every x there is an operation σ of the group with $\sigma a=x$.

If this condition is fulfilled, then, for any two elements x, y, there also exists an operation τ of the group which carries x into y; for from

$$\varrho a=x, \qquad \sigma a=y$$

follows

$$(\sigma\varrho^{-1})x=\sigma a=y.$$

Thus, as regards the question of transitivity it makes no difference from which element a we start.

If the group $\mathfrak{G}$ is not transitive over $\mathfrak{M}$ (*intransitive group*), the set $\mathfrak{M}$ resolves into "*transitivity sets*," i.e., subsets which are transformed into themselves by the groups, and over which the group is transitive. These subsets are obtained according to the following principle: Two elements a, b of $\mathfrak{M}$ shall belong to the same subset if there exists in $\mathfrak{G}$ an operation σ which carries a into b.

This property is 1. reflexive, 2. symmetric, and 3. transitive; for we have

1. $\sigma a=a$ for $\sigma=1$,
2. $\sigma a=b$ implies $\sigma^{-1}b=a$,
3. $\sigma a=b$ and $\tau b=c$ imply $(\tau\sigma)a=c$.

Thus, a partition of $\mathfrak{M}$ into equivalence classes is actually defined.

If a group $\mathfrak{G}$ is transitive over $\mathfrak{M}$, and $\mathfrak{G}_a$ is the subgroup of those elements of $\mathfrak{G}$ which leave fixed the element a of $\mathfrak{M}$, then every left-sided coset $\tau\mathfrak{G}_a$ of $\mathfrak{G}_a$ transforms the element a into the sole element τa. In this manner a one-to-one correspondence between the left-sided cosets and the elements of $\mathfrak{M}$ is obtained, as can readily be seen from the transitivity of $\mathfrak{G}$. The number of the cosets (the index of $\mathfrak{G}_a$) is equal to the number of elements of $\mathfrak{M}$. The group of those elements of $\mathfrak{G}$ which leave invariant another element τa is given by

$$\mathfrak{G}_{\tau a}=\tau\mathfrak{G}_a\tau^{-1}.$$

A transitive group of permutations of a set $\mathfrak{M}$ is called *imprimitive* if it is possible to separate $\mathfrak{M}$ into at least two mutually exclusive subsets $\mathfrak{M}_1, \mathfrak{M}_2, \ldots$ not all of which consist of only one element in such a way that the transformations of

the group carry every set $\mathfrak{M}_\mu$ into a set $\mathfrak{M}_\nu$. The sets $\mathfrak{M}_1, \mathfrak{M}_2, \ldots$ are called *systems of imprimitivity*. If such a separation

$$\mathfrak{M} = \{\mathfrak{M}_1, \mathfrak{M}_2, \ldots\}$$

is impossible, the group is called a *primitive group*.

EXAMPLES. Klein's four-group is imprimitive, the subsets

$$\{1, 2\}, \quad \{3, 4\}$$

being the imprimitive systems. (Two more partitions into systems of imprimitivity are possible.) On the other hand, the complete permutation group (and also the alternating group) of n objects is never primitive; for in every separation of the set $\mathfrak{M}$ into subsets, such as

$$\mathfrak{M} = \{\{1, 2, \ldots, k\}, \{\ldots\}, \ldots\} \qquad (1 < k < n),$$

there exists a permutation which carries $\{1, 2, \ldots, k\}$ into $\{1, 2 \ldots, k-1, k+1\}$, i.e., into a set which is neither disjoint to $\{1, 2, \ldots, k\}$ nor identical with it.

In a separation $\mathfrak{M} = \{\mathfrak{M}_1, \ldots, \mathfrak{M}_r\}$ having the above property, namely that the group $\mathfrak{G}$ permutes the sets $\mathfrak{M}_\nu$ among themselves, there exists for every ν a permutation belonging to the group, which carries $\mathfrak{M}_1$ into $\mathfrak{M}_\nu$. By virtue of the transitivity, we need only find a permutation which carries an arbitrary element of $\mathfrak{M}_1$ into an element of $\mathfrak{M}_\nu$; then this permutation must carry $\mathfrak{M}_1$ into $\mathfrak{M}_\nu$. This implies that every one of the sets $\mathfrak{M}_1, \mathfrak{M}_2, \ldots$ consists of the same number of elements.

For arbitrary transitive permutation groups $\mathfrak{G}$ of a set $\mathfrak{M}$ the following theorem is valid:

Let $\mathfrak{g}$ be the subgroup of those elements of $\mathfrak{G}$ which leave an element a of $\mathfrak{M}$ invariant. If the group $\mathfrak{G}$ is imprimitive, there exists a group $\mathfrak{h}$ distinct from $\mathfrak{g}$ and $\mathfrak{G}$ such that

$$\mathfrak{g} \subset \mathfrak{h} \subset \mathfrak{G},$$

and, conversely, if such an intermediate group $\mathfrak{h}$ exists, $\mathfrak{G}$ is imprimitive. The group $\mathfrak{h}$ leaves a system of imprimitivity $\mathfrak{M}_1$ invariant, and the left-sided cosets of $\mathfrak{h}$ carry $\mathfrak{M}_1$ into the individual systems $\mathfrak{M}_\nu$.

PROOF. First, let $\mathfrak{G}$ be imprimitive, and let $\mathfrak{M} = \{\mathfrak{M}_1, \mathfrak{M}_2, \ldots\}$ be a separation into systems of imprimitivity. Let $\mathfrak{M}_1$ contain the element a. Let $\mathfrak{h}$ be the subgroup of those elements of $\mathfrak{G}$ which leave $\mathfrak{M}_1$ invariant. By the above observation, $\mathfrak{h}$ contains all the permutations of $\mathfrak{G}$ which carry a into itself or into some other element of $\mathfrak{M}_1$; from this follows $\mathfrak{g} \subset \mathfrak{h}$ and $\mathfrak{h} \neq \mathfrak{g}$. But in $\mathfrak{G}$ there exist also permutations which, for example, carry $\mathfrak{M}_1$ into $\mathfrak{M}_2$; hence $\mathfrak{h} \neq \mathfrak{G}$. If, furthermore, τ carries the system $\mathfrak{M}_1$ into $\mathfrak{M}_\nu$, then the entire coset $\tau\mathfrak{h}$ carries $\mathfrak{M}_1$ into $\mathfrak{M}_\nu$.

Conversely, let there be given a group $\mathfrak{h}$ distinct from $\mathfrak{g}$ and $\mathfrak{G}$ such that

$$\mathfrak{g} \subset \mathfrak{h} \subset \mathfrak{G}.$$

$\mathfrak{G}$ completely resolves into cosets $\tau\mathfrak{h}$, and each one of them again resolves into cosets $\sigma\mathfrak{g}$. The latter cosets carry every a into another element σa; thus, if we collect them to cosets $\tau\mathfrak{h}$, then the elements σa are also collected to at least two sets $\mathfrak{M}_1$, $\mathfrak{M}_2$, ... disjoint in pairs, each of the sets consisting of at least two elements. Thus the $\mathfrak{M}_\nu$ are defined as

$$\mathfrak{M}_\nu = \tau\mathfrak{h}a. \tag{1}$$

Any further substitution σ carries $\mathfrak{M}_\nu = \tau\mathfrak{h}a$ into $\sigma\tau\mathfrak{h}a$, i.e., again into a set of the same kind, which proves the imprimitivity of the group. If we denote by $\mathfrak{M}_1$ the set arising from (1) for $\tau = 1$, then $\mathfrak{h}$ leaves fixed the system $\mathfrak{M}_1$, since $\mathfrak{h}\mathfrak{M}_1 = \mathfrak{h}\mathfrak{h}a = \mathfrak{h}a = \mathfrak{M}_1$, and the cosets $\tau\mathfrak{h}$ carry $\mathfrak{M}_1$ into the other systems $\mathfrak{M}_\nu$, since $\tau\mathfrak{h}\mathfrak{M}_1 = \tau\mathfrak{h}\mathfrak{h}a = \tau\mathfrak{h}a$.

EXERCISES. 1. If the number of the elements of the set $\mathfrak{M}$ is a prime number, every transitive group is primitive.

2. The group $\mathfrak{h}$ defined above is transitive over $\mathfrak{M}_1$.

3. Let the set $\mathfrak{M}$ be separated into 3 imprimitive systems of two elements each; let the group $\mathfrak{G}$ be of order 12. What is

a) the index of $\mathfrak{h}$ in $\mathfrak{G}$,

b) the index of $\mathfrak{g}$ in $\mathfrak{h}$,

c) the order of $\mathfrak{g}$?

4. The order of a transitive group of permutations of a finite number of objects is divisible by the number of these objects.

NOTE. The number of the permuted objects is also called the *degree* of the permutation group.

CHAPTER VII

THE GALOIS THEORY

The Galois theory is concerned with the finite separable extensions of a field K, and in particular with its 1-isomorphisms and 1-automorphisms. It establishes a relationship between the extension fields of K, which are contained in a given normal field, and the subgroups of a certain finite group. This theory affords a solution of various questions regarding the solution of algebraic equations.

All fields in this chapter will be commutative. The field K will be called the *rational field.*

50. THE GALOIS GROUP

If the rational field K is given, every finite separable extension field Σ is generated, according to Section 40, by a "primitive element" $\vartheta : \Sigma = \mathsf{K}(\vartheta)$. By Section 38, the number of "relative" isomorphisms (i.e., isomorphisms that leave all elements of K fixed) which Σ possesses in a suitable extension field Ω is equal to the degree n of Σ with respect to K. For this extension field Ω we may choose the decomposition field of the irreducible polynomial $f(x)$ of which ϑ is a root. This decomposition field is the smallest normal field with respect to K which includes Σ or, as we shall say, the *normal field belonging to* Σ. The relative isomorphisms of $\mathsf{K}(\vartheta)$ can be characterized by the fact that they carry the element ϑ into its conjugates $\vartheta_1, \ldots, \vartheta_n$ in Ω[1]. Then every field element $\varphi(\vartheta) = \sum a_\lambda \vartheta^\lambda$ $(a_\lambda \in \mathsf{K})$ goes into $\varphi(\vartheta_\nu) = \sum a_\lambda \vartheta_\nu^\lambda$ and, therefore, instead of speaking of the isomorphism we may speak of the *substitution* $\vartheta \to \vartheta_\nu$.

We have to bear in mind, however, that the use of the elements ϑ and ϑ_ν is only a device for representing the isomorphisms conveniently, and that the *concept* of an isomorphism is wholly independent of the particular choice of ϑ. It is quite possible to construct the isomorphisms by using more than one field generator, as we shall see later.

If Σ is itself a normal field, then all conjugate fields $\mathsf{K}(\vartheta_\nu)$ coincide with Σ.

For first of all, in this case, all ϑ_ν are contained in $\mathsf{K}(\vartheta)$. Furthermore, since

[1] Of course, ϑ is itself among the $\vartheta_1, \ldots, \vartheta_n$.

the $\mathsf{K}(\vartheta_\nu)$ are equivalent to $\mathsf{K}(\vartheta)$, they are normal themselves. Hence ϑ is contained in every $\mathsf{K}(\vartheta_\nu)$.

Converse: *If Σ is identical with all the conjugate fields $\mathsf{K}(\vartheta_\nu)$, then Σ is normal.*

For on this supposition, Σ is equal to the decomposition field $\mathsf{K}(\vartheta_1 \ldots, \vartheta_n)$ of $f(x)$ and therefore normal.

From now on we assume that $\Sigma = \mathsf{K}(\vartheta)$ is a normal field. Under this assumption, the isomorphisms which carry Σ into its conjugate fields $\mathsf{K}(\vartheta_\nu)$ are *automorphisms* of Σ. Evidently, these automorphisms of Σ (which leave all elements of K fixed) form a group of n elements, which is called the *Galois group of Σ with respect to K* or *relative to K*. This group will play the main part in all our further considerations. We denote it by $\mathfrak{G}$. We emphasize once more: *The order of the Galois group is equal to the degree of the field: $n = (\Sigma : \mathsf{K})$.*

Sometimes it is useful to extend the notion of the Galois group to non-normal fields. Every finite extension field Σ' is contained in a smallest normal field Σ generated by the elements of Σ' and their conjugates. Now, by the Galois group of Σ' we always mean the group of Σ.

In order to find the automorphisms it is by no means necessary to find a primitive element ϑ for the field Σ. We may also generate Σ by several successive adjunctions, say $\Sigma = \mathsf{K}(\alpha_1, \ldots, \alpha_m)$, and first find the 1-isomorphisms of $\mathsf{K}(\alpha_1)$ which carry α_1 into its conjugates. Next, we can extend these 1-isomorphisms to the 1-isomorphisms of $\mathsf{K}(\alpha_1, \alpha_2)$, etc.

An important special case is that in which $\alpha_1, \ldots, \alpha_m$ are the roots of an equation $f(x) = 0$ [2]. *By the Galois group of the equation $f(x) = 0$* or *of the polynomial $f(x)$* we mean the Galois group of the decomposition field $\mathsf{K}(\alpha_1, \ldots, \alpha_m)$ of this equation. Every relative automorphism carries the set of the roots into itself, i.e., every automorphism permutes the roots. If this permutation is known, so is the automorphism. For if $\alpha_1, \ldots, \alpha_m$ are carried into $\alpha_1', \ldots, \alpha_m'$ in sequential order, every element of $\mathsf{K}(\alpha_1, \ldots, \alpha_m)$, written as a rational function $\varphi(\alpha_1, \ldots, \alpha_m)$ must go into the corresponding function $\varphi(\alpha_1', \ldots, \alpha_m')$. *Therefore, the Galois group of an equation can be thought of as a group of permutations of the roots.* We always mean this permutation group whenever we speak of the group of the equation.

Let Δ be an "intermediate" field: $\mathsf{K} \subseteq \Delta \subseteq \Sigma$. By a theorem of Section 35, every (relative) isomorphism of Δ which carries Δ into a conjugate field Δ' within Σ can be continued to an isomorphism of Σ, i.e., to an element of the Galois group. From this follows:

Two intermediate fields Δ and Δ' are conjugate with respect to K if, and only if, they can be carried into one another by a substitution of the Galois group.

If we put $\Delta = \mathsf{K}(\alpha)$, the same reasoning yields:

---

[2] $f(x)$ shall be a polynomial without multiple linear factors.

Two elements α, *and* α' *of* Σ *are conjugate with respect to* K *if, and only if, they can be carried into one another by a substitution of the Galois group of* Σ.

The number of the various conjugates of an element α in Σ is equal to the degree of the irreducible equation for α. If this number is equal to 1, then α is the root of a linear equation and, therefore, contained in K. From this follows:

If an element α *of* Σ *"permits" all substitutions of the Galois group of* Σ, *i.e., if it is transformed into itself by all these substitutions,* α *belongs to the rational field* K.

From all these theorems we realize the great significance of the automorphism group for the study of the properties of the field. For the sake of convenience these theorems were stated for finite extension fields, but by "transfinite induction" they can be easily extended to infinite extensions. They are even valid for inseparable extensions if we merely replace the degree of the field by the reduced degree of the field, and if we change the assertion of the last theorem into: "then a power α^{p^l}, where p is the characteristic, belongs to the rational field K." On the other hand, the "Fundamental Theorem of the Galois Theory," to be established in the following section, holds only for finite separable extensions.

The extension field Σ over K is called *Abelian* if the Galois group is Abelian; it is called *cyclic* if the group is cyclic, etc. Similarly, an equation is called *Abelian, cyclic,* or *primitive* if its Galois group is Abelian, cyclic, or primitive (as permutation group of the roots).

A very simple *example* for the Galois groups is given by the Galois fields $GF(p^m)$ (Section 37) if the prime field contained therein is regarded as the rational field. The 1-automorphisms $s(\alpha \to \alpha^p)$, considered in Section 37, and its powers $s^2, s^3, \ldots, s^m = 1$ leave fixed all elements of Π and thus belong to the Galois group. But since the degree of the field is also m, they form the entire group. Hence this group is cyclic of order m.

EXERCISES. 1. Every rational function of the roots of an equation which is carried into itself by the permutations of the Galois group belongs to the rational field, and vice versa.

2. The Galois group of an equation $f(x) = 0$ which has no double root is transitive (Section 49) if, and only if, the equation is irreducible in the rational field.

3. What are the possibilities for the group of an irreducible equation of the third degree?

4. The group of an equation consists of even permutations only as soon as the square root of the discriminant is contained in the rational field. In the other case only half of its permutations are even.

5. Form the Galois groups of the equations

$$x^3 - 2 = 0,$$
$$x^3 + 2x + 1 = 0,$$
$$x^4 - 5x^2 + 6 = 0;$$

also those of the "cyclotomic equations"

$$x^4+x^2+1=0,$$

$$x^4+1=0,$$

all with respect to the field of rational numbers.

51. THE FUNDAMENTAL THEOREM OF THE GALOIS THEORY

The "Fundamental Theorem" reads as follows:

1. *Every intermediate field* Δ, *where* $K \subseteqq \Delta \subseteqq \Sigma$ *defines a subgroup* $\mathfrak{g}$ *of the Galois group* $\mathfrak{G}$, *namely the totality of those automorphisms of* Σ *which leave fixed all the elements of* Δ. 2. Δ *is uniquely determined by* $\mathfrak{g}$; *for* Δ *is the totality of those elements of* Σ *which "permit" the substitutions of* $\mathfrak{g}$, *i.e., which are left invariant under them.* 3. *For every subgroup* $\mathfrak{g}$ *of* $\mathfrak{G}$ *a field* Δ *can be found which bears the mentioned relationship to* $\mathfrak{g}$. 4. *The order of* $\mathfrak{g}$ *is equal to the degree of* Σ *with respect to* Δ; *the index of* $\mathfrak{g}$ *in* $\mathfrak{G}$ *is equal to the degree of* Δ *with respect to* K.

PROOF. The totality of automorphisms of Σ which leave fixed all elements of Δ is the Galois group of Σ relative to Δ, and thus, in any event, has the group property. This proves proposition 1., while 2. follows from the last theorem of Section 50 applied to Σ, considering Δ as the rational field. Proposition 3. is somewhat more difficult.

Let again $\Sigma = K(\vartheta)$, and let $\mathfrak{g}$ be a given subgroup of $\mathfrak{G}$. By Δ we denote the totality of elements of Σ which are transformed into themselves under the substitutions σ of $\mathfrak{g}$. This Δ, obviously, is a field; for if α and β are left fixed under the substitutions σ, then the same is true for $\alpha+\beta$, $\alpha-\beta$, $\alpha\cdot\beta$, and in case $\beta \neq 0$ also for $\alpha:\beta$. Moreover, we have $K \subseteqq \Delta \subseteqq \Sigma$. The Galois group of Σ relative to Δ includes the group $\mathfrak{g}$, since the substitutions of $\mathfrak{g}$ surely have the property to leave the elements of Δ fixed. If the Galois group of Σ with respect to Δ had more elements than $\mathfrak{g}$ has, then the degree $(\Sigma:\Delta)$ would also be greater than the order of $\mathfrak{g}$. This degree $(\Sigma:\Delta)$ is equal to the degree of ϑ with respect to Δ, since $\Sigma=\Delta(\vartheta)$. If $\sigma_1, \ldots, \sigma_h$ are the substitutions of $\mathfrak{g}$, then ϑ is a root of an equation of the h-th degree

$$(1) \qquad (x-\sigma_1\vartheta)(x-\sigma_2\vartheta)\cdots(x-\sigma_h\vartheta)=0.$$

The coefficients of this equation permit the group $\mathfrak{g}$ and, therefore, belong to Δ. Hence the degree of ϑ with respect to Δ is no greater than the order of $\mathfrak{g}$.

Consequently, there remains only the possibility that $\mathfrak{g}$ is exactly the Galois group of Σ with respect to Δ. This completes the proof of 3. (In addition, the irreducibility of (1) in $\Delta[x]$ follows.)

Finally, if n is the order of $\mathfrak{G}$, h again the order of $\mathfrak{g}$, and j the index, we have

$$n=(\Sigma:\mathsf{K}),\qquad h=(\Sigma:\Delta)_{\mathfrak{i}}\qquad n=h\cdot j,$$
$$(\Sigma:\mathsf{K})=(\Sigma:\Delta)\cdot(\Delta:\mathsf{K}),$$

so that

$$(\Delta:\mathsf{K})=j.$$

This proves 4.

By the Fundamental Theorem just proved, the relationship between the subgroups $\mathfrak{g}$ and the intermediate fields Δ is one-to-one. Now the question arises: How can we find $\mathfrak{g}$ when we have Δ, and how Δ when we have $\mathfrak{g}$?

The first part of the question is easy to answer. Suppose we have found the elements conjugate to ϑ, namely $\vartheta_1,\ldots,\vartheta_n$, all expressed in terms of ϑ; then we also have the automorphisms $\vartheta\to\vartheta_\nu$ which constitute the group $\mathfrak{G}$. Now, if a subfield $\Delta=\mathsf{K}(\beta_1,\ldots,\beta_k)$ is given, where $\beta_1,\ldots,\beta_k$ are known expressions of ϑ, then $\mathfrak{g}$ simply consists of those substitutions of $\mathfrak{G}$ which leave $\beta_1,\ldots,\beta_k$ invariant; for they leave invariant all rational functions of $\beta_1,\ldots,\beta_k$ as well.

Conversely, if $\mathfrak{g}$ is given, we form the product

$$(x-\sigma_1\vartheta)\,(x-\sigma_2\vartheta)\cdots(x-\sigma_k\vartheta).$$

According to the proof of the Fundamental Theorem, the coefficients of this polynomial must lie in Δ and even generate Δ; for they generate a field relative to which the element ϑ, as the root of equation (1), is already of degree h and which, therefore, cannot be a proper subfield of Δ. Thus the generators of Δ are simply the elementary symmetric functions of $\sigma_1\vartheta,\ldots,\sigma_k\vartheta$.

Another method consists in finding a function $\chi(\vartheta)$ which permits the substitutions of $\mathfrak{g}$, but which does not permit any more substitutions of $\mathfrak{G}$. Then the element $\chi(\vartheta)$ will belong to the field Δ, but not to any proper subfield of Δ and will therefore generate Δ. The existence of such a function follows, e.g., from the Theorem of the Primitive Element (Section 40).

Once the Galois group is known, the Fundamental Theorem of the Galois Theory allows us to determine all intermediate fields between K and Σ. Their number is obviously finite; for a finite group has only a finite number of subgroups. The groups also reveal how the various fields are nested in each other; for the following theorem holds:

If Δ_1 is a subfield of Δ_2, then the group $\mathfrak{g}_2$ belonging to Δ_2 is a subgroup of the group $\mathfrak{g}_1$ belonging to Δ_1, and vice versa.

PROOF. Suppose first: $\Delta_1\subseteqq\Delta_2$. Then any substitution which leaves fixed all elements of Δ_2 will leave fixed all elements of Δ_1.

Secondly, let $\mathfrak{g}_1\supseteqq\mathfrak{g}_2$. Then any field element that permits all substitutions of $\mathfrak{g}_1$ will permit all substitutions of $\mathfrak{g}_2$.

Let us finally raise the question: What will happen to the Galois group of $K(\vartheta)$ relative to K when the rational field K is extended to a field Λ and, accordingly, the normal field $K(\vartheta)$ to $\Lambda(\vartheta)$? (Of course, we assume that $\Lambda(\vartheta)$ has a meaning, i.e., that Λ and ϑ are contained in a common extension field Ω.)

The substitutions $\vartheta \to \vartheta_\nu$, which, after the extension, yield automorphisms of $\Lambda(\vartheta)$ will yield isomorphisms of $K(\vartheta)$ as well, and, therefore, since $K(\vartheta)$ is normal, automorphisms of $K(\vartheta)$. *Therefore, after the extension of the rational field, the substitution group is a subgroup of the original one.* That the subgroup may be proper is seen at once by choosing Λ as an intermediate field between K and $K(\vartheta)$. However, it may happen that the subgroup coincides with the original one; in this case we say that the extension of the rational field *does not reduce* the group of $K(\vartheta)$.

EXERCISES. 1. The union field of the fields belonging to two subgroups of the Galois group $\mathfrak{G}$ belongs to the intersection of these two subgroups, and the intersection fields belongs to the union group.[3]

2. If the field Σ is, with respect to K, cyclic of degree n, then for every divisor d of n there exists exactly one intermediate field Δ of degree d, and two such intermediate fields are contained in one another when, and only when, the degree of one of them is divisible by the degree of the other (cf. Section 8, Ex. 6).

3. With the aid of the Galois theory, determine the subfields of $GF(p^n)$ anew (Section 37).

4. Let $K \subseteq \Lambda$ and $K(\vartheta)$ be normal over K. Show that the group of $K(\vartheta)$ relative to K is equal to that of $\Lambda(\vartheta)$ relative to Λ if, and only if, $K(\vartheta) \cap \Lambda = K$.

5. Prove by means of the theorem of Section 49:

The field $K(\alpha_1)$ which arises by the adjunction of a root of an irreducible algebraic equation possesses a subfield Δ such that

$$K \subset \Delta \subset K(\alpha_1)$$

if, and only if, the Galois group of the equation, as a permutation group of the roots, is imprimitive. In particular, Δ can be so determined that the degree of the field (Δ/K) is equal to the number of systems of imprimitivity, and that the equation in Δ is decomposed into irreducible factors which correspond to the systems of imprimitivity.

6. Show that the Fundamental Theorem also holds for inseparable extensions (characteristic p) subject only to the following modifications. Proposition 2 becomes: The totality of the elements of Σ which permit the substitution of $\mathfrak{g}$ is the "root field of Δ in Σ," i.e., the totality of the elements of Σ, a p^l-th power of which belongs to Δ. Proposition 3 becomes: To every subgroup of $\mathfrak{g}$ we can find exactly one field Δ which is invariant under the operation of extracting the

[3] By the union group of two subgroups is meant the group generated by the union of two sets; the term union field is defined in a similar way.

p-th root, and which permits the substitutions of $\mathfrak{g}$ and no others. Proposition 4 is valid for the reduced degrees.

52. CONJUGATE GROUPS, CONJUGATE FIELDS AND ELEMENTS

Again, let $\mathfrak{G}$ be the Galois group of Σ relative to K, and let β be an element of Σ. The subgroup $\mathfrak{g}$ which belongs to the intermediate field $\mathsf{K}(\beta)$ consists of the substitutions which leave β invariant. The other substitutions of $\mathfrak{G}$ transform β into the quantities conjugate to it, and every conjugate quantity can be obtained in this way (Section 50). Moreover, we assert:

The substitutions of $\mathfrak{G}$ which transform β into a given conjugate element form a coset $\tau\mathfrak{g}$ of $\mathfrak{g}$, and every coset transforms β into a single conjugate element.

PROOF. If ϱ and τ are substitutions which carry β into the same conjugate element

$$\varrho(\beta) = \tau(\beta),$$

it follows that

$$\tau^{-1}\varrho(\beta) = \tau^{-1}\tau(\beta) = \beta;$$

hence $\tau^{-1}\varrho = \sigma$ is an element of $\mathfrak{g}$, and we have $\varrho = \tau\sigma$; thus ϱ and τ lie in the same coset $\tau\mathfrak{g}$. If, conversely, ϱ and τ lie in the same coset, i.e., if both lie in $\tau\mathfrak{g}$, we have $\varrho = \tau\sigma$, where σ lies in $\mathfrak{g}$; therefore we have

$$\varrho(\beta) = \tau\sigma(\beta) = \tau(\sigma(\beta)) = \tau(\beta).$$

From this theorem follows anew that the degree of β (= number of the conjugates) is equal to the index of $\mathfrak{g}$ (= number of the cosets).

An automorphism τ which carries β into $\tau\beta$ carries $\mathsf{K}(\beta)$ into the conjugate field $\mathsf{K}(\tau\beta)$. We assert: *The field $\mathsf{K}(\tau\beta)$ belongs to the subgroup $\tau\mathfrak{g}\tau^{-1}$.*

For the subgroup belonging to $\mathsf{K}(\tau\beta)$ consists of the substitutions σ' which leave $\tau\beta$ invariant so that

$$\sigma'\tau\beta = \tau\beta,$$

or

$$\tau^{-1}\sigma'\tau\beta = \beta,$$

or

$$\tau^{-1}\sigma'\tau = \sigma \qquad \text{in } \mathfrak{g},$$

or

$$\sigma' = \tau\sigma\tau^{-1},$$

i.e., we are dealing exactly with the group $\tau\mathfrak{g}\tau^{-1}$.

Consequently, to conjugate fields belong conjugate groups. By Section 50, a field Δ over K is normal when, and only when, it is identical with all its conjugate fields. From this follows:

A field Δ, $K \leqq \Delta \leqq \Sigma$, *is normal if, and only if, the group* $\mathfrak{g}$ *is identical with all its conjugates* $\tau\mathfrak{g}\tau^{-1}$ *in* $\mathfrak{G}$, *i.e., if it is a normal divisor in* $\mathfrak{G}$.

Now, if Δ is normal, the following question arises at once: What is the group of Δ relative to K?

Every automorphism in $\mathfrak{G}$ transforms Δ into itself and thus induces an automorphism of Δ belonging to the required group of Δ over K. To the product of two automorphisms in $\mathfrak{G}$ corresponds the product of the induced automorphisms of Δ so that $\mathfrak{G}$ is homomorphically mapped upon the group of Δ. The elements in $\mathfrak{G}$, to which corresponds the identity substitution of Δ, are exactly those of $\mathfrak{g}$; from this follows by the Law of Homomorphism (Section 10) that the required group is 1-isomorphic with the factor group $\mathfrak{G}/\mathfrak{g}$. Hence we have:

The Galois group of Δ *relative to* K *is isomorphic with the factor group* $\mathfrak{G}/\mathfrak{g}$.

EXERCISES. 1. All subfields of an Abelian field are Galois and are themselves Abelian. All subfields of a cyclic field are themselves cyclic.

2. If $K \leqq \Delta \leqq \Sigma$, and if Δ is the smallest normal field over K which includes Δ, then the group belonging to Δ is the intersection of the group belonging to Δ and its conjugate groups.

3. What are the subfields of the field $\Gamma(\varrho, \sqrt[3]{2})$, where Γ is the field of rational numbers, and $\varrho = \frac{-1-\sqrt{-3}}{2}$ a primitive third root of unity? What are the degrees of the fields? Which subfields are conjugates, which are normal?

4. Answer the same questions for the field $\Gamma(\sqrt{2}, \sqrt{5})$.

53. CYCLOTOMIC FIELDS

Let Γ be the field of rationals, i.e., the prime field of characteristic zero. Consider the equation which has as its roots the primitive h-th roots of unity, each counted once. This equation

$$\Phi_h(x) = 0 \tag{1}$$

(cf. Section 36) is called the *cyclotomic equation*, and the field of the h-th roots of unity is called a *cyclotomic field*. The reason for this is the following: The complex number

$$\zeta = e^{\frac{2\pi i}{h}} = \cos\frac{2\pi}{h} + i\sin\frac{2\pi}{h}$$

is a primitive h-th root of unity, from which $\cos\frac{2\pi}{h}$ can be determined by the equation

$$2\cos\frac{2\pi}{h} = \zeta + \zeta^{-1},$$

and knowing this cosine, we can construct a regular polygon of h sides, i.e., we can divide the circle into h equal arcs.

The following theory of the cyclotomic fields is valid, regardless whether the primitive root of unity ζ is viewed as a complex number, or just as a mere symbol.

First of all, we have to show that equation (1) is irreducible in Γ.

Let the irreducible equation which is satisfied by an arbitrarily chosen primitive root of unity ζ be $f(\zeta)=0$. The polynomial $f(x)$ may be assumed to be a primitive polynomial with integer coefficients. We have to show that $f(x)=\Phi_h(x)$.

Let p be a prime number which does not divide h. Then ζ is, like ζ^p, a primitive h-th root of unity, and satisfies a primitive irreducible equation $g(\zeta^p)=0$ with integer coefficients. First we want to show that $f(x)=\varepsilon g(x)$, where $\varepsilon=\pm 1$ is a unit in the ring of integers.

The polynomial x^h-1 has the zero ζ in common with $f(x)$, and the zero ζ^p with $g(x)$, so it is divisible by $f(x)$ as well as by $g(x)$. If $f(x)$ and $g(x)$ were essentially different (i.e., if they would differ not only by one unit as factor) x^h-1 would have to be divisible by $f(x)g(x)$,

$$x^h-1=f(x)\,g(x)\,h(x), \tag{2}$$

where, by Section 23, $h(x)$ is again a polynomial with integer coefficients. Moreover, the polynomial $g(x^p)$ has the root ζ and must therefore be divisible by $f(x)$:

$$g(x^p)=f(x)\,k(x). \tag{3}$$

Again $k(x)$ is a polynomial with integral coefficients.

We now regard (2) and (3) as congruences modulo p. We have modulo p:

$$g(x^p)\equiv\{g(x)\}^p.$$

For we can perform the exponentiation on the right by first writing $g(x)$ as a sum of powers of x without coefficients (by replacing, e.g., $2x^3$ by x^3+x^3) and next, by raising each individual term to the p-th power (cf. Section 30, Ex. 2). Doing this, we obtain exactly $g(x^p)$. Hence (3) implies

$$\{g(x)\}^p\equiv f(x)\,k(x) \pmod p \tag{4}$$

Let us now suppose both sides of (4) are resolved into irreducible factors (mod p). By the unique factorization theorem for polynomials with coefficients in the field $C/(p)$ (cf. Section 17), an arbitrary prime factor $\varphi(x)$ of $f(x)$ must also occur in $\{g(x)\}^p$, and therefore in $g(x)$. The right member of (2) is therefore divisible by $\varphi(x)^2$ modulo p so that the left member x^h-1 and its derivative hx^{h-1} must both be divisible by $\varphi(x)$ modulo p But because of $h\not\equiv 0 \pmod p$, hx^{h-1} has only prime factors x, which do not divide x^h-1. Thus we have been led to a contradiction.

Therefore, we have indeed $f(x)=\pm g(x)$, and ζ^p is a root of $f(x)$.

Next we show: All primitive roots of unity are roots of $f(x)$. Let ζ^ν be a primitive root of unity and

$$\nu=p_1\cdots p_n,$$

where the p_i are identical or distinct prime factors, but, at any rate, relatively prime to h.

Since ζ satisfies the equation $f(x)=0$, so must ζ^{p_1} by what has just been proved. By repeating the conclusion for the prime number p_2, we see that $\zeta^{p_1 p_2}$ satisfies the equation, too. Continuing in this way, we find (by complete induction) that ζ^r satisfies the equation $f(x)=0$.

Thus all zeros of $\Phi_h(x)$ satisfy the equation $f(x)=0$. Since $f(x)$ was irreducible, and since $\Phi_h(x)$ has no multiple factors, it follows that

$$\Phi_h(x)=f(x).$$

This proves the *irreducibility of the cyclotomic equation.*[4]

On the basis of this fact alone we can without difficulty construct the Galois group of the cyclotomic field $\Gamma(\zeta)$.

In the first place the degree of the field is equal to the degree of $\Phi_h(x)$, and hence equal to $\varphi(h)$ (cf. Section 36). An automorphism of $\Gamma(\zeta)$ transforms ζ into some other root of $\Phi_h(x)$. All powers ζ^λ, where λ is relatively prime to h, are roots of $\Phi_h(x)$. Let σ_λ be the automorphism which carries ζ into ζ^λ. We have

$$\sigma_\lambda=\sigma_\mu$$

as soon as

$$\zeta^\lambda=\zeta^\mu,$$

which means

$$\lambda\equiv\mu\,(h).$$

Moreover, we have

$$\sigma_\lambda\sigma_\mu(\zeta)=\sigma_\lambda(\zeta^\mu)=\{\sigma_\lambda(\zeta)\}^\mu=\zeta^{\lambda\mu}$$

so that

$$\sigma_\lambda\sigma_\mu=\sigma_{\lambda\mu}.$$

The automorphism group of $\Gamma(\zeta)$ is therefore isomorphic with the group of those residue classes mod h which are relatively prime to h (cf. Section 19, Ex. 6).

This group is Abelian. Consequently, all subgroups, normal divisors, and all subfields are normal and Abelian.

EXAMPLE: The twelfth roots of unity. The residue classes relatively prime to 12 are represented by

$$1,\ 5,\ 7,\ 11$$

The automorphisms may therefore be denoted by σ_1, σ_5, σ_7, σ_{11}. The automorphism σ_λ carries ζ into ζ^λ. The multiplication table reads:

σ_1	σ_5	σ_7	σ_{11}
σ_5	σ_1	σ_{11}	σ_7
σ_7	σ_{11}	σ_1	σ_5
σ_{11}	σ_7	σ_5	σ_1

[4] Other simple proofs may be found in two successive articles by E. Landau and I. Schur in *Math. Z.* Volume 29 (1929).

Every element is of order 2. Thus, besides the group itself and the identity group, there are just three subgroups:

1. $\{\sigma_1, \sigma_5\}$,
2. $\{\sigma_1, \sigma_7\}$,
3. $\{\sigma_1, \sigma_{11}\}$.

To these groups belong quadratic fields generated by square roots. We may find the latter as follows:

The fourth roots of unity i, $-i$ are also twelfth roots of unity, and therefore lie in the field. Therefore $\Gamma(i)$ is a quadratic subfield.

Similarly, the third roots of unity lie in the field. Since

$$\varrho = -\tfrac{1}{2} + \tfrac{1}{2}\sqrt{-3}$$

is a third root of unity, $\Gamma\left(\sqrt{-3}\right)$ is a quadratic subfield.

By multiplying i by $\sqrt{-3}$, we get $\sqrt{3}$. Hence $\Gamma\left(\sqrt{3}\right)$ is the third subfield.

We now ask which subgroups belong to these three fields.

Since $\sigma_5\zeta^3 = \zeta^{15} = \zeta^3$, the element $i = \zeta^3$ permits the automorphism σ_5. Hence $\Gamma(i)$ belongs to the group $\{\sigma_1, \sigma_5\}$.

$\varrho = \zeta^4$ permits the automorphism σ_7, since $\sigma_7\zeta^4 = \zeta^{28} = \zeta^4$. Hence $\Gamma\left(\sqrt{-3}\right)$ belongs to the group $\{\sigma_1, \sigma_7\}$.

The remaining field $\Gamma\left(\sqrt{3}\right)$ must belong to the group $\{\sigma_1, \sigma_{11}\}$.

Any two of the three subfields generate the whole field. Therefore it must be possible to express the root of unity ζ by means of two square roots. In fact we have

$$\zeta = \zeta^{-3}\zeta^4 = i^{-1}\varrho = -i\frac{-1+\sqrt{-3}}{2} = \frac{i-\sqrt{3}}{2}.$$

In the following section we shall learn how to determine explicity the subfields for cyclotomic fields with prime exponents, and how to construct the cyclotomic field itself from these subfields by successive adjunctions.

EXERCISES. 1. For $h > 2$ the quantity $\zeta + \zeta^{-1}$ always generates a subfield of degree $\frac{1}{2}\varphi(h)$.

2. Find the group and the subfields of the field of the fifth roots of unity, and express the latter by square roots. The same for the eighth roots of unity.

3. Find the group and the subfields of the field of the seventh roots of unity. What is the defining equation of the field $\Gamma(\zeta + \zeta^{-1})$?

54. THE PERIODS OF THE CYCLOTOMIC EQUATION

Let the exponent h of the roots of unity under consideration now be a prime number q. In this case the cyclotomic equation reads:

$$\Phi_q(x) = \frac{x^q - 1}{x - 1} = x^{q-1} + x^{q-2} + \cdots + x + 1 = 0.$$

It is of degree $n = q - 1$.

Let ζ be a primitive q-th root of unity.

The group of the residue classes relative prime to q is cyclic (Section 37), and therefore consists of the n residue classes

$$1,\ g,\ g^2,\ \ldots,\ g^{n-1},$$

where g is a "primitive number mod q" or a primitive root of the congruence $g^n \equiv 1\,(q)$. *Therefore, the Galois group is likewise cyclic* and is generated by that automorphism σ which carries ζ into ζ^g. The primitive roots of unity may be represented as follows:

$$\zeta,\ \zeta^g,\ \zeta^{g^2},\ \ldots,\ \zeta^{g^{n-1}},\quad \text{where } \zeta^{g^n}=\zeta.$$

We put

$$\zeta^{g^\nu}=\zeta_\nu.$$

Since

$$\zeta^{g^{\nu+n}}=\zeta^{g^\nu},$$

we can operate modulo n with the numbers ν.

We have

$$\sigma(\zeta_i)=\sigma(\zeta^{g^i})=\{\sigma(\zeta)\}^{g^i}=(\zeta^g)^{g^i}=\zeta^{g^{i+1}}=\zeta_{i+1}.$$

Thus the automorphism σ raises every index by 1. The ν-fold repetition of σ yields

$$\sigma^\nu(\zeta_i)=\zeta_{i+\nu}.$$

The $\zeta_i (i=0,1,\ldots,n-1)$ form a *field basis.* In order to recognize this fact we only have to show that they are linearly independent. As a matter of fact, the ζ_i coincide with the $\zeta,\ldots\zeta^{q-1}$, except for the order in which they occur; thus, a linear relation among them would mean:

$$a_1\zeta+\cdots+a_{q-1}\zeta^{q-1}=0,$$

or, after factoring out a factor ζ,

$$a_1+a_2\zeta+\cdots+a_{q-1}\zeta^{q-2}=0.$$

Since ζ cannot satisfy any equation of degree $\leqq q-2$, this implies

$$a_1=a_2=\cdots=a_{q-1}=0;$$

hence the ζ_i are linearly independent.

The subfields of the cyclotomic field are obtained at once from the subgroups of the cyclic group (cf. end of Section 7):

If

$$ef=n$$

is a decomposition of n into two positive factors, there exists a subgroup $\mathfrak{g}$ *of order f which consists of the elements*

$$\sigma^e,\ \sigma^{2e},\ \ldots,\ \sigma^{(f-1)e},\ \sigma^{fe},$$

where σ^{fe} *is the identity. Any subgroup can be obtained in this manner.*

We now look for the elements α which permit σ^e (and hence permit the subgroup $\mathfrak{g}$). If

$$\alpha = a_0\zeta_0 + \cdots + a_{n-1}\zeta_{n-1},$$

then

$$\sigma^e(\alpha) = a_0\zeta_e + a_1\zeta_{e+1} + \cdots + a_{n-1}\zeta_{e+n-1}. \tag{1}$$

If this is to be equal to α, we must have

$$\begin{aligned} a_0 &= a_e, \\ &\cdots\cdots \\ a_\nu &= a_{e+\nu}, \\ &\cdots\cdots \\ a_{n-1} &= a_{e+n-1}, \end{aligned}$$

where the indices are to be taken modulo n. It follows that

$$a_\nu = a_{\nu+e} = a_{\nu+2e} = \cdots;$$

so that we can collect the terms of (1) to form groups

$$a_\nu(\zeta_\nu + \zeta_{\nu+e} + \cdots).$$

Therefore, we put

$$\eta_\nu = \zeta_\nu + \zeta_{\nu+e} + \zeta_{\nu+2e} + \cdots + \zeta_{\nu+(f-1)e} \quad (\nu = 0, \ldots, e-1) \tag{2}$$

and write for (1)

$$\alpha = a_0\eta_0 + a_1\eta_1 + \cdots + a_{e-1}\eta_{e-1}.$$

From this we see that the η_i form a basis for the subfield belonging to $\mathfrak{g}$.

We have

$$\begin{aligned} \sigma(\eta_0) &= \eta_1, \\ &\cdots\cdots \\ \sigma^\nu(\eta_0) &= \eta_\nu, \end{aligned}$$

so that $\eta_0, \eta_1, \ldots$ are conjugate, and the polynomial

$$(x-\eta_0)(x-\eta_1)\cdots(x-\eta_{e-1}) \tag{3}$$

is irreducible.

Since the field $\Gamma(\eta_0)$, like any subfield, is normal, the polynomial (3) is decomposed in it completely; hence the field under consideration is already generated by η_0:

$$\Gamma(\eta_0) = \Gamma(\eta_0, \ldots, \eta_{e-1}).$$

The quantities $\eta_0, \ldots, \eta_{e-1}$ defined by (2) are called, according to Gauss, the periods of the cyclotomic field.

Gauss established a formula by means of which a product $\eta_i\eta_k$ can be computed conveniently. He introduced the new notation

$$\begin{aligned} \eta^{(r)} &= \zeta^r + \zeta^{rg^e} + \cdots + \zeta^{rg^{(f-1)e}} \\ &= \sum_{\nu \bmod f} \zeta^{rg^{\nu e}} \end{aligned}$$

(where the notation "$\nu \bmod f$" means that ν runs over a system of representatives of the residue classes modulo f). Thus, for $r \not\equiv 0 \pmod q$, $\eta^{(r)}$ is that η_ν in which a term ζ^r occurs. We notice that

$$\eta^{(r g^{\varepsilon})} = \eta^{(r)},$$

and that

$$\eta^{(0)} = 1 + \cdots + 1 = f.$$

Multiplication of two $\eta^{(r)}$ yields

$$\eta^{(r)}\eta^{(s)} = \sum_{\nu \bmod f} \left(\sum_{\mu \bmod f} \zeta^{r g^{\nu e} + s g^{\mu e}} \right),$$

or for $\mu = \mu' + \nu$:

$$\eta^{(r)}\eta^{(s)} = \sum_{\nu \bmod f} \left(\sum_{\mu' \bmod f} \zeta^{r g^{\nu e} + s g^{(\mu' + \nu) e}} \right)$$
$$= \sum_{\mu' \bmod} \left(\sum_{\nu \bmod f} \zeta^{(r g^{\nu e} + s g^{(\mu' + \nu) e}} \right).$$

The quantity in parentheses is $\eta^{(r + s g^{\mu' e})}$; thus, when we write μ again instead of μ', it follows that

$$\eta^{(r)}\eta^{(s)} = \sum_{\mu \bmod f} \eta^{(r + s g^{\mu e})} \qquad \text{(Gauss' formula).}$$

The indices of the η in this sum coincide with the exponents, which are obtained by multiplying the first term of $\eta^{(r)}$ by all terms of $\eta^{(s)}$.

From

$$\zeta + \zeta^2 + \cdots + \zeta^{q-1} = -1$$

follows

$$\eta_0 + \eta_1 + \cdots + \eta_{e-1} = -1$$

so that

$$\eta^{(0)} = f = -f(\eta_0 + \cdots + \eta_{e-1}).$$

Thus every $\eta^{(0)}$ on the right of Gauss' formula may be removed. Noting that the other $\eta^{(j)}$, except for their order, coincide with $\eta_0, \ldots, \eta_{e-1}$, we obtain for every product $\eta_i \eta_k$ a representation as the sum of integral multiples of the η_i.

EXAMPLE: $q = 17$. The number 3 is a primitive root; for the powers of 3 modulo 17 are the following:

Exponents:	0	1	2	3	4	5	6	7	8
Powers:	1	3	−8	−7	−4	5	−2	−6	−1
Exponents:	9	10	11	12	13	14	15	16	
Powers:	−3	8	7	4	−5	2	6	1	

We first compute *the periods of 8 terms* ($e = 2$, $f = 8$):

$$\eta_0 = \zeta + \zeta^{-8} + \zeta^{-4} + \zeta^{-2} + \zeta^{-1} + \zeta^{8} + \zeta^{4} + \zeta^{2},$$
$$\eta_1 = \zeta^{3} + \zeta^{-7} + \zeta^{5} + \zeta^{-6} + \zeta^{-3} + \zeta^{7} + \zeta^{-5} + \zeta^{6}.$$

We have $\eta_0 + \eta_1 = -1$, and by Gauss' formula (since $\eta_0 = \eta^{(1)}$, $\eta_1 = \eta^{(3)}$):

$$\eta_0\eta_1 = \eta^{(4)} + \eta^{(-6)} + \eta^{(6)} + \eta^{(-5)} + \eta^{(-2)} + \eta^{(8)} + \eta^{(-4)} + \eta^{(7)}.$$

Now $\eta^{(r)}$ is always that η_ν in which ζ^r occurs. Hence

$$\eta^{(4)} = \eta^{(-2)} = \eta^{(8)} = \eta^{(-4)} = \eta_0,$$

and

$$\eta^{(-6)} = \eta^{(6)} = \eta^{(-5)} = \eta^{(7)} = \eta_1$$

so that

$$\eta_0\eta_1 = 4\eta_0 + 4\eta_1 = -4.$$

Therefore, η_0 and η_1 are the roots of the equation

(4) $$y^2 + y - 4 = 0.$$

The solutions of this equation are

$$y = -\frac{1}{2} \pm \frac{1}{2}\sqrt{17}.$$

The *periods of 4 terms* $(e = 4,\ f = 4)$ are:

$$\xi_0 = \zeta + \zeta^{-4} + \zeta^{-1} + \zeta^4,$$
$$\xi_1 = \zeta^3 + \zeta^5 + \zeta^{-3} + \zeta^{-5},$$
$$\xi_2 = \zeta^{-8} + \zeta^{-2} + \zeta^8 + \zeta^2,$$
$$\xi_3 = \zeta^{-7} + \zeta^{-6} + \zeta^7 + \zeta^6.$$

We have

$$\xi_0 + \xi_2 = \eta_0,$$
$$\xi_1 + \xi_3 = \eta_1.$$

In order to find an equation for ξ_0 and ξ_2 we compute

$$\begin{aligned}\xi_0\xi_2 &= \xi^{(-7)} + \xi^{(-1)} + \xi^{(-8)} + \xi^{(3)}\\ &= \xi_3 + \xi_0 + \xi_2 + \xi_1\\ &= -1.\end{aligned}$$

Hence ξ_0 and ξ_2 satisfy the equation

(5) $$x^2 - \eta_0 x - 1 = 0.$$

Similarly, ξ_1 and ξ_3 satisfy the equation

(6) $$x^2 - \eta_1 x - 1 = 0.$$

These equations express the fact known beforehand that $\Gamma(\xi_0)$ is quadratic with respect to $\Gamma(\eta_0)$.

Two *periods of 2 terms* are

$$\lambda^{(1)} = \zeta + \zeta^{-1},$$
$$\lambda^{(4)} = \zeta^4 + \zeta^{-4}.$$

Addition and multiplication yield:

$$\lambda^{(1)} + \lambda^{(4)} = \xi_0,$$
$$\lambda^{(1)}\lambda^{(4)} = \zeta^5 + \zeta^{-3} + \zeta^3 + \zeta^{-5} = \xi_1.$$

Therefore $\lambda^{(1)}$ and $\lambda^{(4)}$ satisfy the equation

$$\Lambda^2 - \xi_0\Lambda + \xi_1 = 0.$$

Finally, ζ itself satisfies the equation

$$\zeta + \zeta^{-1} = \lambda^{(1)},$$

or

$$\zeta^2 - \lambda^{(1)}\zeta + 1 = 0.$$

Thus the 17th roots of unity are expressed by roots of quadratic equations.

If we consider the 17th roots of unity as complex numbers, we can put

$$\zeta = e^{\frac{2\pi i}{17}},$$

$$\lambda^{(1)} = \zeta + \zeta^{-1} = 2\cos\frac{2\pi}{17}.$$

Equation (4) has a positive and a negative root. Since we have

$$\eta_0 = (\zeta + \zeta^{-1}) + (\zeta^8 + \zeta^{-8}) + (\zeta^4 + \zeta^{-4}) + (\zeta^2 + \zeta^{-2})$$
$$= 2\left(\cos\frac{2\pi}{17} + \cos\frac{16\pi}{17} + \cos\frac{8\pi}{17} + \cos\frac{4\pi}{17}\right) > 2\left(\frac{1}{2} - 1 + 0 + \frac{1}{2}\right) = 0,$$

η_0 is the positive root:

$$\eta_0 = -\frac{1}{2} + \frac{1}{2}\sqrt{17}.$$

Similarly, (5) and (6) have a positive and negative root each. Since

$$\xi_0 = 2\left(\cos\frac{2\pi}{17} + \cos\frac{8\pi}{17}\right) > 0,$$

$$\xi_3 = 2\left(\cos\frac{14\pi}{17} + \cos\frac{12\pi}{17}\right) < 0,$$

we have ξ_0 and ξ_1 as the positive roots of (5) and (6). Finally

$$\lambda^{(1)} = 2\cos\frac{2\pi}{17} > 2\cos\frac{8\pi}{17} = \lambda^{(4)}$$

is the larger of the two (positive) roots of (7). With the aid of these formulae we can *construct a regular polygon of* 17 *sides* with ruler and compass (cf. Section 59).

EXERCISES. 1. Construct a regular polygon of 17 sides.

2. For the periods η_0 and η_1 of $\frac{p-1}{2}$ terms prove the general relations

$$\eta_0 + \eta_1 = -1$$

$$\eta_0\eta_1 = \frac{1+p}{4} \text{ for } p \equiv -1 \ (4),$$

$$\eta_0\eta_1 = \frac{1-p}{4} \text{ for } p \equiv 1 \ (4),$$

and derive a quadratic equation for η_0.

3. The η_0 of Ex. 2 is the "Gaussian sum":

$$\eta_0 = \sum_{s=1}^{\frac{p-1}{2}} \zeta^{s^2}.$$

55. CYCLIC FIELDS AND PURE EQUATIONS

In this section we assume that the rational field K contains the n-th roots of unity, and that n times the identity is not zero (i.e. n is not divisible by the characteristic). Under these assumptions the following proposition holds: *The group of a "pure equation"*

$$x^n - a = 0 \qquad (a \neq 0)$$

relative to K *is cyclic.*

PROOF. If ϑ is one root of the equation, then $\zeta\vartheta, \zeta^2\vartheta, \ldots, \zeta^{n-1}\vartheta$ (where ζ is a primitive n-th root of unity) are the others.[5] Therefore, ϑ already generates the field of the roots, and every substitution of the Galois group is of the form

$$\vartheta \to \zeta^\nu \vartheta.$$

The composition of two substitutions $\vartheta \to \zeta^\nu \vartheta$ and $\vartheta \to \zeta^\mu \vartheta$ yields $\vartheta \to \zeta^{\mu+\nu}\vartheta$. Thus a definite root of unity ζ^ν corresponds to every substitution, and the product of the roots of unity corresponds to the product of the substitutions. Therefore, the Galois group is isomorphic with a subgroup of the group of the n-th roots of unity. Since the latter group is cyclic, each of its subgroups, and therefore the Galois group, is cyclic.

If, in particular, the equation $x^n - a = 0$ is irreducible, all roots $\zeta^\nu\vartheta$ are conjugate to ϑ, and therefore the Galois group is isomorphic with the entire group of the n-th roots of unity. In this case its order is n.

Next we show that, conversely, every cyclic field of the n-th degree over K can be generated by roots of pure equations $x^n - a = 0$.

Let $\Sigma = \mathsf{K}(\vartheta)$ be a cyclic field of degree n, and let σ be the generating substitution of the Galois group so that $\sigma^n = 1$. Again we assume that the rational field K contains the n-th roots of unity.

If ζ is such an n-th root of unity, we form the "*Lagrange resolvent*":

$$(\zeta, \vartheta) = \vartheta_0 + \zeta\vartheta_1 + \cdots + \zeta^{n-1}\vartheta_{n-1}, \tag{1}$$

where

$$\vartheta_\nu = \sigma^\nu \vartheta.$$

By the substitution σ the ϑ_ν undergo a cyclic permutation

$$\sigma\vartheta_\nu = \vartheta_{\nu+1} \qquad (\vartheta_n = \vartheta_0),$$

and the resolvent (ζ, ϑ) goes into

$$\begin{aligned}\sigma(\zeta, \vartheta) &= \vartheta_1 + \zeta\vartheta_2 + \cdots + \zeta^{n-2}\vartheta_{n-1} + \zeta^{n-1}\vartheta_0 \\ &= \zeta^{-1}(\vartheta_0 + \zeta\vartheta_1 + \zeta^2\vartheta_2 + \cdots + \zeta^{n-1}\vartheta_{n-1}) \\ &= \zeta^{-1}(\zeta, \vartheta).\end{aligned}$$

Hence the n-th power $(\zeta, \vartheta)^n$ remains unaltered in the substitution σ, i.e., $(\zeta, \vartheta)^n$ belongs to the rational field K.

We can obtain $(\zeta, \vartheta)^n$ as a purely formal expression from (1) by exponentiation and find

$$(\zeta, \vartheta)^n = P_0 + \zeta P_1 + \cdots + \zeta^{n-1} P_{n-1}, \tag{2}$$

where the P_ν are polynomials of degree n in the ϑ which are independent of the root of unity ζ employed at the outset.

If we multiply (1) by $\zeta^{-\nu}$ and perform a summation over all ζ, we obtain (noting the last theorem of Section 36):

$$\sum_\zeta \zeta^{-\nu}(\zeta, \vartheta) = n\vartheta_\nu. \tag{3}$$

[5] Evidently, all the roots are different so that the equation is separable.

Since, by hypothesis, the number n is not divisible by the characteristic of the field, we can compute ϑ_r from (3), provided the (ζ, ϑ) are known. But because of (2), each (ζ, ϑ) is an n-th root of an element of the rational field K. From this we obtain the desired result:

Every cyclic field of degree n can be generated by the adjunction of n-th roots, provided that the rational field contains the n-th roots of unity, and that n is not divisible by the characteristic.

It is useful to note that $(\zeta, \vartheta) \cdot (\zeta^{-1}, \vartheta)$ is not altered by the substitution σ, since the first factor is multiplied by ζ^{-1}, and the second by ζ. Consequently

$$(\zeta, \vartheta) \cdot (\zeta^{-1}, \vartheta)$$

belongs to the rational field. Therefore, only one of every two such "conjugate" resolvents need be adjoined.

Finally

$$(1, \vartheta) = \vartheta_0 + \vartheta_1 + \cdots + \vartheta_{n-1}$$

belongs to K, as is readily seen.

If our field Σ arises by the adjunction of the roots $\xi_1, \ldots, \xi_m$ of an equation $f(x) = 0$, then σ effects a permutation of these roots and, therefore, a permutation of their subscripts $1, 2, \ldots, m$. Let this permutation be resolved into cycles:

$$(1\ 2 \ldots j)\,(j+1 \ldots l) \ldots$$

The permutations of the Galois group are the powers of the one written out. They carry the subscript 1 into $1, 2, 3, \ldots, j$. Suppose now the equation $f(x) = 0$ is irreducible; then all roots are conjugate; therefore, it must be possible to carry the root ξ_1 into all other roots, i.e., the one cycle $(1\ 2 \ldots j)$ must contain all roots. Since the generating permutation of the cycle must be of order n, we must have $j = n$. Thus, the degree m of the equation is likewise equal to n and therefore equal to the degree of the field; hence the adjunction of one root must already generate the entire field. If we now number the roots by $0, 1, \ldots, n-1$ instead of $1, 2, \ldots n$, we can choose our field generator $\vartheta = \vartheta_0$ to be ξ_0; by numbering the rest of the roots in a suitable way, we automatically obtain $\vartheta_1 = \sigma\vartheta = \sigma\xi_0 = \xi_1$, $\vartheta_2 = \sigma\vartheta_1 = \sigma\xi_1 = \xi_2$, etc.

Therefore, the roots of $f(x)$, when suitably numbered, can be chosen as the ϑ, in (1).

If the rational field K does not contain the n-th roots of unity, then, in order to be able to apply the above method of solution by means of n-th roots, we first have to adjoin the n-th roots of unity ζ to K. In this adjunction the Galois group remains cyclic, since a subgroup of a cyclic group is always cyclic.

We proceed to furnish some criteria for the *irreducibility of the pure equations of prime degree p.*

First, if the rational field K again contains the p-th roots of unity, then, by what was proved at the beginning of this section, the group is a subgroup of a cyclic

group of order p. So it is either the complete group or the identity group. In the first case all the roots are conjugate, and the equation is irreducible. In the second case all roots are invariant under the substitutions of the Galois group, and the equation resolves into linear factors in the field K. Therefore: *The polynomial $x^p - a$ either resolves completely, or it is irreducible.*

If K does not contain the roots of unity, we are unable to assert as much as that. But the following theorem is valid:

Either $x^p - a$ is irreducible, or a is a p-th power in K, so that there exists in K a decomposition

$$x^p - a = x^p - \beta^p = (x-\beta)(x^{p-1} + \beta x^{p-2} + \cdots + \beta^{p-1}).$$

PROOF. Let us suppose $x^p - a$ is reducible:

$$x^p - a = \varphi(x) \cdot \psi(x).$$

In its decomposition field $x^p - a$ resolves as follows:

$$x^p - a = \prod_{\nu=0}^{p-1} (x - \zeta^\nu \vartheta) \qquad (\vartheta^p = a).$$

Therefore, the factor $\varphi(x)$ must be a product of certain factors $x - \zeta^\nu \vartheta$, and the constant term $\pm b$ of $\varphi(x)$ must have the form $\pm \zeta' \vartheta^\mu$, where ζ' is a p-th root of unity:

$$b = \zeta' \vartheta^\mu,$$
$$b^p = \vartheta^{p\mu} = a^\mu.$$

Because of $0 < \mu < p$ we have $(\mu, p) = 1$, and so, with suitable rational integers ϱ and σ, we have

$$\varrho\mu + \sigma p = 1,$$
$$a = a^{\varrho\mu} a^{\sigma p} = b^{\varrho p} a^{\sigma p};$$

hence a is a p-th power.

Interesting theorems on the reducibility of pure equations may be found in papers by A. Cappelli: "Sulla riducibilità dell'equazioni algebriche," *Rendiconti Napoli* 1898, and by G. Darbi: "Sulla riducibilità dell'equazioni algebriche." *Annali di Mat.* (4) 4 (1926).

EXERCISE. 1. If we drop the assumption that the rational field K contains the n-th roots of unity, the group of the pure equation $x^n - a = 0$ is isomorphic with a group of linear substitutions modulo n:

$$x' \equiv c x + b.$$

[The normal field obtained by adjoining all the roots to K is $K(\vartheta, \zeta)$, and for every substitution σ of the group we have

$$\sigma\zeta = \zeta^c,$$
$$\sigma\vartheta = \zeta^b \vartheta].$$

56. SOLUTION OF EQUATIONS BY RADICALS

As is well known, the roots of an equation of the second, third or fourth degrees can be found from the coefficients by rational operations and by extractions of roots $\sqrt{\ }, \sqrt[3]{\ }, \ldots$ ("radicals") (cf. Section 58). We now raise the question as to what equations have the property that their roots can be expressed in terms of the elements of the "rational" field K through rational operations and radicals. Since reducible equations can be decomposed into prime factors, we may limit ourselves to irreducible equations with coefficients in K. The problem consists in constructing over K a field containing one or all roots of the given equation, by successive adjunctions of radicals $\sqrt[n]{a}$ such that a belongs to the field already constructed each time.

In one point, however, the statement of the problem is still too vague. In general, the radical sign $\sqrt[n]{\ }$ in a field is a many-valued function, and the question is which root is meant by $\sqrt[n]{a}$ each time. For example, if we express a primitive sixth root of unity by radicals by simply representing them by $\sqrt[6]{1}$, or even by $\sqrt[12]{1}$, this must be regarded as an unsatisfactory solution, whereas the solution $\zeta = \frac{1}{2} \pm \frac{1}{2}\sqrt{-3}$ is much more satisfactory, because the expression $\frac{1}{2} \pm \frac{1}{2}\sqrt{-3}$ represents just the two primitive sixth roots of unity for *any* choice of the value of $\sqrt{-3}$ (i.e., of a solution of the equation $x^2+3=0$).

The most rigid requirement in this respect is that, first, *all* solutions of the equation in question are to be represented by expressions of the form

$$\sqrt[n]{\cdots \sqrt[m]{\cdots} + \sqrt[r]{\cdots} + \cdots + \cdots} \tag{1}$$

(or similar ones) and, secondly, that these expressions are to represent solutions of the equation of *any* choice of the radicals appearing in these expressions. Of course, if a radical $\sqrt[m]{a}$ occurs several times in expression (1), the same value has to be assigned to it every time it occurs.

Suppose the first requirement is fulfilled. Then the second will also be fulfilled, provided we can see to it that in the successive adjunction of the radicals $\sqrt[n]{a}$ the respective equation $x^n - a = 0$ is always *irreducible*, as each adjunction is performed. For in this case all possible choices of the $\sqrt[n]{a}$ will always yield conjugate quantities, which can be carried into each other by isomorphisms, and in all further adjunctions these isomorphisms can be continued to isomorphisms of the extension fields (cf. Section 35). Thus if, in determining the values of the radicals $\sqrt[n]{a}$ the expression (1) constitutes a root of the equation in question, it must constitute a root in every determination of the value of the root of the equation in question, since every isomorphism always carries the roots of a polynomial in $\mathsf{K}[x]$ into the same roots.

After these preliminary remarks we are in a position to formulate the Fundamental Theorem on Equations Solvable by Radicals:

1. *If one root of an equation* $f(x)=0$ *irreducible in* K *can be represented by an expression* (1), *and if the radical exponents are not divisible* [6] *by the characteristic of the field* K, *then the group of this equation is soluble* (*i.e., its composition factors are cyclic of prime order*). 2. *If, conversely, the group of the equation is soluble, then all roots can be represented by expressions* (1) *in such a way that in the successive adjunctions of the* $\sqrt[n]{a}$ *the exponents are prime numbers and the equations* $x^n-a=0$ *are irreducible each time, provided the characteristic of the field* K *is zero or larger than the largest prime number which occurs among the orders of the composition factors.*[7]

Essentially, this theorem states that the solubility of the group is decisive for the solubility of the equation by radicals. The concept of solubility by radicals is expressed as weakly as possible in the first part of the theorem, while in the second part it is expressed as strongly as possible so that the theorem asserts as much as possible.

PROOF. 1. First, we can make all radical exponents in (1) prime numbers by writing

$$\sqrt[rs]{a}=\sqrt[r]{\sqrt[s]{a}}.$$

Then we adjoin to K all p_1-th, p_2-th, etc. roots of unity, where $p_1, p_2, \ldots$ are the prime numbers occurring in (1) as radical exponents. These adjunctions give rise to a series of successive cyclic normal extensions, which may be decomposed into extensions of prime degree. But as soon as these roots of unity are in the field, the adjunction of $\sqrt[p]{a}$ is, by Section 55, either no extension at all, or it is a cyclic normal extension of degree p. Now, as soon as we have adjoined a $\sqrt[p]{a}$, we successively adjoin all p-th roots of the elements conjugate to a; these adjunctions are either no extensions at all or cyclic extensions of prime degree, and by these adjunctions our fields always remain normal with respect to K. Thus, by a series of cyclic adjunctions

$$\mathsf{K}\subset\Lambda_1\subset\Lambda_2\subset\cdots\subset\Lambda_\omega, \tag{2}$$

we finally arrive at a normal field $\Lambda_\omega=\Omega$ which contains the expression (1), a root of $f(x)$. Since the field Ω is normal, it contains all roots of $f(x)$, i.e., it contains the decomposition field Σ of $f(x)$.

Let $\mathfrak{G}$ be the Galois group of Ω with respect to K. To the chain of fields (2) there corresponds a chain of subgroups of $\mathfrak{G}$:

$$\mathfrak{G}\supset\mathfrak{G}_1\supset\mathfrak{G}_2\supset\cdots\supset\mathfrak{G}_\omega=\mathfrak{E}, \tag{3}$$

[6] The purpose of this assumption is to prevent the appearance of inseparable extensions. We could drop it; but here we are not interested in that.

[7] If we admit, besides radicals of the kind described, roots of unity in the solution formula, then the last requirement may be replaced by a weaker one: among the orders of the composition factors the characteristic shall not occur.

and each of these groups is normal divisor in the preceding one, the factor group being cyclic of prime order. This means that the group $\mathfrak{G}$ is soluble, and that (3) is a composition series.

To the field Σ belongs a subgroup $\mathfrak{H}$, which is a normal divisor of $\mathfrak{G}$, and by Section 46 we can lay a composition series through $\mathfrak{H}$ which, except for isomorphism, has the same composition factors, but may possibly have a different order:

$$\mathfrak{G} \supset \mathfrak{H}_1 \supset \mathfrak{H}_2 \supset \cdots \supset \mathfrak{H} \supset \cdots \supset \mathfrak{E}. \tag{4}$$

The Galois group of Σ relative to K is the group $\mathfrak{G}/\mathfrak{H}$, for which we now have the composition series

$$\mathfrak{G}/\mathfrak{H} \supset \mathfrak{H}_1/\mathfrak{H} \supset \mathfrak{H}_2/\mathfrak{H} \supset \cdots \supset \mathfrak{H}/\mathfrak{H} = \mathfrak{E}.$$

By the second law of isomorphism (Section 45), the factors in this series are 1-isomorphic with the respective factors of (4), and so again cyclic of prime order. This proves proposition 1.

Regarding proposition 2, we first prove the following

LEMMA. *The q-th roots of unity (where q is a prime) are expressible as "irreducible radicals" (i.e., roots of irreducible equations $x^p - a = 0$), provided the characteristic of K is zero or larger than q.*

Since the proposition is trivial for $q = 2$ (the second roots of unity ± 1 are rational), we may assume it to be proved for all prime numbers smaller than q. The field of the q-th roots of unity is cyclic of degree $q - 1$, and when we decompose $q - 1$ into primes $q - 1 = p_1^{e_1} \dots p_r^{e_r}$, we can construct this field by a sequence of cyclic extensions of degree p_ν. If we adjoin beforehand the p_1-th, . . . , p_r-th roots of unity, which, by the induction hypothesis, are expressible as radicals, then we can employ for the cyclic extensions of the degrees p_ν the theorem of Section 55, according to which successive field generators are expressible as radicals. The respective equations $x^{p_\nu} - a = 0$ must be irreducible, since otherwise the degree of the fields could not be equal to the p_ν.

Now we can prove proposition 2. Let Σ be the decomposition field of $f(x)$, and let $\mathfrak{G} \supset \mathfrak{G}_1 \supset \cdots \supset \mathfrak{G}_l = \mathfrak{E}$ be a composition series for the Galois group of Σ with respect to K. To this series of groups belongs a series of fields:

$$\mathsf{K} \subset \Lambda_1 \subset \cdots \subset \mathsf{K}_l = \Sigma,$$

each of which is normal and cyclic with respect to its predecessor. If $q_1, q_2, \dots$ are the relative degrees occurring in the series, we adjoin to K first the q_1-st, q_2-nd, etc. roots of unity. According to the lemma, this is possible by irreducible radicals. Then, by the theorem of Section 55, the generators of $\Lambda_1, \Lambda_2, \dots, \Lambda_l$ can be expressed as radicals, the respective equations $x^{q_\nu} - a = 0$ being either irreducible, or they resolve completely each time (end of Section 55). In the latter case the adjunction of the respective radicals is redundant. Thus we have proved proposition 2.

The following example will show that proposition 2 is actually false if one of the degrees q is equal to the characteristic p of the field. The "general equation of the second degree" x^2+ux+v (where u, v are indeterminates which are adjoined to the prime field of characteristic 2) is irreducible and separable and remains irreducible upon adjunction of all roots of unity. The adjunction of a root of an irreducible pure equation of odd degree fails to decompose the equation, since this adjunction produces a field of odd degree. Similarly, the adjunction of a square root fails to decompose the equation, since this adjunction does not change the reduced degree of the field. Therefore, there is no way to solve the equation by radicals.

AN APPLICATION. The symmetric permutation groups 2, 3, or 4 digits (and their subgroups) are soluble; this explains the possibility of the solution of equations of the second, third, and fourth degrees by radicals (see Section 58). The symmetric group of 5 or more digits, however, is no longer soluble (Section 48), and we shall presently see that there are equations of any degree having actually the symmetric group; therefore, there are no general solution formulae for the equations of the fifth or higher degrees. Only special kinds of such equations (like the cyclotomic equations) can be solved by radicals.

Fields or equations with soluble groups are called *metacyclic*. Sometimes the group itself is called metacyclic (instead of soluble).

57. THE GENERIC EQUATION OF DEGREE n

By the *generic equation of degree n* we mean the equation

$$z^n-u_1z^{n-1}+u_2z^{n-2}-+\cdots+(-1)^nu_n=0, \tag{1}$$

with *indeterminate* coefficients $u_1,\ldots,u_n$, which are adjoined to the rational field K. If their roots are $v_1,\ldots,v_n$, we have

$$\begin{cases} u_1=v_1+\cdots+v_n, \\ u_2=v_1v_2+v_1v_3+\cdots+v_{n-1}v_n, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ u_n=v_1v_2\ldots v_n. \end{cases}$$

We compare the generic equation with another one whose roots are indeterminates $x_1,\ldots,x_n$, and whose coefficients are therefore the elementary symmetric functions of these indeterminates:

$$\begin{cases} z^n-\sigma_1z^{n-1}+\sigma_2z^{n-2}-+\cdots+(-1)^n\sigma_n \\ \quad=(z-x_1)(z-x_2)\cdots(z-x_n)=0; \end{cases} \tag{2}$$

$$\begin{cases} \sigma_1=x_1+\cdots+x_n, \\ \sigma_2=x_1x_2+x_1x_3+\cdots+x_{n-1}x_n, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \sigma_n=x_1x_2\ldots x_n. \end{cases}$$

Equation (2) is separable, and its Galois group with respect to the field $\mathsf{K}(\sigma_1, \ldots, \sigma_n)$ is the symmetric group of all the permutations of the x_ν; for every such permutation constitutes a 1-automorphism of the field $\mathsf{K}(x_1, \ldots, x_n)$, which leaves invariant the symmetric functions $\sigma_1, \ldots, \sigma_n$ and, therefore, all elements of the field $\mathsf{K}(\sigma_1, \ldots, \sigma_n)$. Any function of the $x_1, \ldots, x_n$ that is left invariant under the permutations of the group thus belongs to the field $\mathsf{K}(\sigma_1, \ldots \sigma_n)$, i.e., *every symmetric function of the x_ν can be expressed rationally in terms of $\sigma_1, \ldots \sigma_n$.* Thus, by means of the Galois theory, we have once more proved the main part of the "Fundamental Theorem on Symmetric Functions" in Section 26.

We also obtain without difficulty the "Uniqueness Theorem" of Section 26 once more, i.e., *the fact that no relation $f(\sigma_1, \ldots, \sigma_n) = 0$ can exist unless the polynomial f itself vanishes identically.* For suppose we had

$$f(\sigma_1, \ldots, \sigma_n) = f(\Sigma x_i, \Sigma x_i x_k, \ldots, x_1 x_2 \ldots x_n) = 0,$$

this relation would subsist if we substitute the v for the indeterminates x_i. We would therefore have

$$f(\Sigma v_i, \Sigma v_i v_k, \ldots, v_1 v_2 \ldots v_n) = 0,$$

or $f(u_1, \ldots, u_n) = 0$; hence f would vanish identically.

It follows from the Uniqueness Theorem that the correspondence

$$f(u_1, \ldots, u_n) \to f(\sigma_1, \ldots, \sigma_n)$$

is not only a homomorphism, but a 1-isomorphism of the rings $\mathsf{K}[u_1, \ldots, u_n]$ and $\mathsf{K}[\sigma_1, \ldots, \sigma_n]$. It is possible to extend it to an isomorphism of the quotient fields $\mathsf{K}(u_1, \ldots, u_n)$ and $\mathsf{K}(\sigma_1, \ldots, \sigma_n)$ and, moreover, to an isomorphism of the splitting fields $\mathsf{K}(v_1, \ldots, v_n)$ and $\mathsf{K}(x_1, \ldots, x_n)$, according to Section 35. The v_i go into the x_k in some sequential order; since the x_k, however, are permutable, we can let every v_i go into x_i. Thus we have proved:

There exists an isomorphism

$$\mathsf{K}(v_1, \ldots, v_n) \cong \mathsf{K}(x_1, \ldots, x_n),$$

which carries every v_i into x_i, and every u_i into σ_i.

By virtue of this isomorphism all theorems concerning equation (2) can immediately be applied to (1). In particular, we obtain:

The generic equation (1) *is separable, and its Galois group with respect to its coefficient field $\mathsf{K}(u_1, \ldots, u_n)$ is the symmetric group.* The degree of its decomposition field is $n!$.

We put

$$\mathsf{K}(u_1, \ldots, u_n) = \Delta,$$
$$\mathsf{K}(v_1, \ldots, v_n) = \Sigma$$

and denote the symmetric group by $\mathfrak{S}_n$. It always possesses a subgroup of index 2, the alternating group $\mathfrak{A}_n$. The intermediate field Δ belonging to it is of degree 2 and it is generated by any function of the v_i which admits $\mathfrak{A}_n$, but not $\mathfrak{S}_n$. If

the characteristic of K is distinct from zero, such a function is given by the *difference product*

$$\prod_{i<k}(v_i - v_k) = \sqrt{D}.$$

The square of this product is the *discriminant* of equation (1):

$$D = \prod_{i<k}(v_i - v_k)^2.$$

The discriminant is a symmetric function, and hence a polynomial in the u_i. Thus the field Λ is obtained in the form

$$\Lambda = \Delta(\sqrt{D}).$$

For $n > 4$ the group $\mathfrak{A}_n$ is simple (Section 48); hence

$$\mathfrak{S}_n > \mathfrak{A}_n > \mathfrak{E} \tag{3}$$

is a composition series. Thus, for $n > 4$ the group $\mathfrak{S}_n$ is not soluble, and from this follows, by Section 56, the famous theorem first proved by Abel:

The generic equation of degree n is not soluble by radicals for $n > 4$.

For $n = 2$ and $n = 3$ the composition factors in (3) are cyclic. For $n = 2$ we have even $\mathfrak{A}_n = \mathfrak{E}$; for $n = 3$ the factors are of orders 2 and 3. For $n = 4$ we have the composition series

$$\mathfrak{S}_n > \mathfrak{A}_n > \mathfrak{B}_4 > \mathfrak{Z}_2 > \mathfrak{Z},$$

where $\mathfrak{B}_4$ is "the four-group" (Vierergruppe)

$$\{1, \quad (1\,2)\ (3\,4), \quad (1\,3)\ (2\,4), \quad (1\,4)\ (2\,3)\},$$

and $\mathfrak{Z}_2$ any one of its subgroups of order 2. The composition factors are of orders

$$2, 3, 2, 2.$$

On these facts rest the solution formulae of the equation of the second, third, and fourth degrees, which we shall treat in the following section.

58. EQUATIONS OF THE SECOND, THIRD, AND FOURTH DEGREES

According to the general theory, the solution of the generic *equation of the second degree*

$$x^2 + px + q = 0$$

must be possible by means of one square root, for which we can choose (cf. the end of the preceding section) the difference product of the roots x_1, x_2:

$$x_1 - x_2 = \sqrt{D}; \qquad D = p^2 - 4q.$$

From this and from

$$x_1 + x_2 = -p$$

we obtain the well-known solution formulae

$$x_1 = \frac{-p+\sqrt{D}}{2}, \qquad x_2 = \frac{-p-\sqrt{D}}{2},$$

provided that the characteristic of the field is not 2.

The generic *equation of the third degree*

$$z^3 + a_1 z^2 + a_2 z + a_3 = 0$$

can be written in the form [8]

$$x^3 + p x + q = 0$$

by substituting

$$z = x - \tfrac{1}{3} a_1.$$

(In accordance with the general theory of solution in Section 56, we assume that the characteristic of the rational field is distinct from 2 and 3.)

According to the composition series

$$\mathfrak{S}_3 > \mathfrak{A}_3 > \mathfrak{E}$$

we first adjoin the difference product of the roots:

$$(x_1 - x_2)(x_1 - x_3)(x_2 - x_3) = \sqrt{D} = \sqrt{-4p^3 - 27q^2},$$

(cf. end of Section 26, where we have to set $a_1 = 0$, $a_2 = p$, $a_3 = -q$). This adjunction gives rise to a field $\Delta(\sqrt{D})$, relative to which the equation has the group $\mathfrak{A}_3$, i.e., a cyclic group of order 3. According to the general theory of Section 55, we first adjoin the third roots of unity,

$$\varrho = -\tfrac{1}{2} + \tfrac{1}{2}\sqrt{-3}, \qquad \varrho^2 = -\tfrac{1}{2} - \tfrac{1}{2}\sqrt{-3}, \tag{1}$$

and, next, consider Lagrange's resolvents:

$$\begin{aligned} (1, x_1) &= x_1 + x_2 + x_3 = 0, \\ (\varrho, x_1) &= x_1 + \varrho x_2 + \varrho^2 x_3, \\ (\varrho^2, x_1) &= x_1 + \varrho^2 x_2 + \varrho x_3. \end{aligned}$$

The third power of each of these quantities must be a rational expression in $\sqrt{-3}$ and $\sqrt{D}$. The calculation yields:

$$\begin{aligned} (\varrho, x_1)^3 = {} & x_1^3 + x_2^3 + x_3^3 \\ & + 3\varrho x_1^2 x_2 + 3\varrho x_2^2 x_3 + 3\varrho x_3^2 x_1 \\ & + 3\varrho^2 x_1 x_2^2 + 3\varrho^2 x_2 x_3^2 + 3\varrho^2 x_3 x_1^2 \\ & + 6 x_1 x_2 x_3. \end{aligned}$$

By interchanging ϱ and ϱ^2 we obtain $(\varrho^2, x_1)^3$. By substituting (1) in it and noting that

$$\begin{aligned} \sqrt{D} &= (x_1 - x_2)(x_1 - x_3)(x_2 - x_3) \\ &= x_1^2 x_2 + x_2^2 x_3 + x_3^2 x_1 - x_1 x_3^2 - x_2 x_3^2 - x_3 x_1^2, \end{aligned}$$

---

[8] Just for the purpose of simplifying the formulae. The same proof readily yields the solution formulae for the original equation

$$z^3 + a_1 z^2 + a_2 z + a_3 = 0.$$

we get

$$(\varrho, x_1)^3 = \sum x_1^3 - \tfrac{3}{2}\sum x_1^2 x_2 + 6 x_1 x_2 x_3 + \tfrac{3}{2}\sqrt{-3}\sqrt{D}.$$

(The meaning of the sums is the same as for symmetric functions, Section 26.) The symmetric functions which are involved here may easily be expressed, by Section 26, in terms of the elementary symmetric functions $\sigma_1, \sigma_2, \sigma_3$ and therefore in terms of the coefficients of our equation. We have

$$\begin{aligned}
\sigma_1^3 &= \sum x_1^3 + 3\sum x_1^2 x_2 + 6 x_1 x_2 x_3 \quad = 0 \text{ since } \sigma_1 = 0,\\
-\tfrac{9}{2}\sigma_1\sigma_2 &= -\tfrac{9}{2}\sum x_1^2 x_2 - \tfrac{27}{2} x_1 x_2 x_3 = 0 \text{ since } \cdot\sigma_1 = 0,\\
\tfrac{27}{2}\sigma_3 &= \tfrac{27}{2} x_1 x_2 x_3 = -\tfrac{27}{2} q\\
&\overline{\sum x_1^3 - \tfrac{3}{2}\sum x_1^2 x_2 + 6 x_1 x_2 x_3 = -\tfrac{27}{2} q;}
\end{aligned}$$

hence

$$(\varrho, x_1)^3 = -\tfrac{27}{2} q + \tfrac{3}{2}\sqrt{-3}\sqrt{D},$$

and similarly

$$(\varrho^2, x_1)^3 = -\tfrac{27}{2} q - \tfrac{3}{2}\sqrt{-3}\sqrt{D}.$$

The two cubic irrationalities (ϱ, x_1) and (ϱ^2, x_1) are not independent, but we have (cf. Section 55)

$$\begin{aligned}
(\varrho, x_1)\cdot(\varrho^2, x_1) &= x_1^3 + x_2^3 + x_3^3 + (\varrho+\varrho^2) x_1 x_2 + (\varrho+\varrho^2) x_1 x_3 + (\varrho+\varrho^2) x_2 x_3\\
&= x_1^2 + x_2^2 + x_3^2 - x_1 x_2 - x_1 x_3 - x_2 x_3\\
&= \sigma_1^2 - 3\sigma_2 = -3\rho.
\end{aligned}$$

Thus the cubic roots

$$(2)\qquad (\varrho, x_1) = \sqrt[3]{-\tfrac{27}{2} q + \tfrac{3}{2}\sqrt{-3D}}, \qquad (\varrho^2, x_1) = \sqrt[3]{-\tfrac{27}{2} q - \tfrac{3}{2}\sqrt{-3D}}$$

have to be determined in such manner that their product becomes

$$(3)\qquad (\varrho, x_1)\cdot(\varrho^2, x_1) = -3\rho.$$

The roots x_1, x_2, x_3 are obtained by aid of equation (3) (Section 55) as follows:

$$(4)\qquad \begin{cases} 3\cdot x_1 = \sum\limits_{\zeta} (\zeta, x_1) = (\varrho, x_1) + (\varrho^2, x_1),\\ 3\cdot x_2 = \sum\limits_{\zeta} \zeta^{-1} (\zeta, x_1) = \varrho^2 (\varrho, x_1) + \varrho (\varrho^2, x_1),\\ 3\cdot x_3 = \sum\limits_{\zeta} \zeta^{-2} (\zeta, x_1) = \varrho (\varrho, x_1) + \varrho^2 (\varrho^2, x_1). \end{cases}$$

Formulae (2), (3), (4) contain "*Cardano's solution.*" By virtue of their derivation they hold not only for the "generic," but for any special cubic equation as well.

CONCERNING REALITY. If the rational field to which the coefficients p, q belong is a real number field K, two cases are possible:

a) The equation has a real and two conjugate-complex roots. Then, obviously,

$$(x_1-x_2)\,(x_1-x_3)\,(x_2-x_3)$$

is purely imaginary, so that $D<0$. The quantities $\pm\sqrt{-3D}$ are real, and for (ϱ, x_1) in (2) we can choose a real cube root. Because of (3), (ϱ^2, x_1) will be real as well, and formula (4) gives $3x_1$ as the sum of two real cube roots, while x_2 and x_3 are represented by (4) as conjugate-complex quantities.

b) The equation has three real roots. In this case $\sqrt{D}$ is real, so that $D\geqq 0$. In case $D=0$ (two roots are equal) the case is the same as before; for $D>0$, however, the expressions under the cube root sign in (2) become imaginary, and so the (real) expressions (4) are obtained as the sums of *imaginary* cube roots, i.e., not in a real form.

This is the so-called "*casus irreducibilis*" of the cubic equation. We shall show that *in this case it is actually impossible to solve the equation*

$$x^3+px+q=0$$

by real radicals unless the equation resolves in the rational field K *already.*

Let the equation $x^3+px+q=0$ be irreducible in K, and let it have three real roots x_1, x_2, x_3. First adjoin $\sqrt{D}$. This adjunction does not make the equation reducible [since the at most quadratic field $\mathsf{K}(\sqrt{D})$ cannot contain a root of an irreducible cubic equation], and its group now becomes $\mathfrak{A}_3$. Suppose now that the equation could be decomposed by a number of adjunctions of real radicals. The radical exponents may of course be assumed to be prime numbers. Under these assumptions there would exist among these adjunctions a "critical" adjunction $\sqrt[h]{a}$ (where h is prime) which would effect the decomposition, while before this adjunction, say in the field Δ, the equation would still be irreducible. By Section 55, either x^h-a is irreducible in Δ, or a is a h-th power of a number in Δ. In the second case the real h-th root in a would be contained in Δ already so that its adjunction could not effect the decomposition. Therefore we must assume that x^h-a is irreducible, and that the degree of the field $\Delta(\sqrt[h]{a})$ is exactly h. By hypothesis, a root of the equation $x^3+px+q=0$, which is still irreducible in Δ, is contained in $\Delta(\sqrt[h]{a})$; hence h is divisible by 3. This implies $h=3$, and we have, say $\Delta(\sqrt[3]{a})=\Delta(x_1)$. The decomposition field $\Delta(x_1, x_2, x_3)$ relative to Δ is also of degree 3; hence $\Delta(\sqrt[3]{a})=\Delta(x_1, x_2, x_3)$. The field $\Delta(\sqrt[3]{a})$ which has now been identified as a normal field must contain besides $\sqrt[3]{a}$ the conjugate elements $\varrho\sqrt[3]{a}$ and $\varrho^2\sqrt[3]{a}$ and therefore the roots of unity ϱ and ϱ^2. Thus we have been led to a contradiction; for the field $\Delta(\sqrt[3]{a})$ is real, and the number ϱ is not.

Just so, the generic *equation of the fourth degree*

$$z^4+a_1z^3+a_2z^2+a_3z+a_4=0$$

can be transformed into

$$x^4+px^2+qx+r=0$$

by the substitution

$$z=x-\tfrac{1}{4}a_1.$$

To the composition series

$$\mathfrak{S}_4>\mathfrak{A}_4>\mathfrak{B}_4>\mathfrak{Z}_2>\mathfrak{E}$$

belongs a series of fields

$$\Delta<\Delta(\sqrt{D})<\Delta_1<\Delta_2<\Sigma.$$

Once more, let the characteristic of Δ be $\neq 2$ and $\neq 3$. As we shall see, it is not necessary to determine D explicitly. The field Δ_1 is generated from $\Delta(\sqrt{D})$ by

an element which permits the substitutions of $\mathfrak{B}_4$, but not of $\mathfrak{A}_4$; such an element is

$$\Theta_1 = (x_1 + x_2)(x_3 + x_4).$$

This element, incidentally, permits also the following substitutions besides those of $\mathfrak{B}_4$:

$$(1\ 2),\quad (3\ 4),\quad (1\ 3\ 2\ 4),\quad (1\ 4\ 2\ 3).$$

These substitutions, together with $\mathfrak{B}_4$, form a group of order 8. The generating element has three different conjugates with respect to Δ, into which it is carried by the substitutions of $\mathfrak{S}_4$, viz.:

$$\begin{aligned}\Theta_1 &= (x_1 + x_2)(x_3 + x_4),\\ \Theta_2 &= (x_1 + x_3)(x_2 + x_4),\\ \Theta_3 &= (x_1 + x_4)(x_2 + x_3).\end{aligned}$$

These conjugates are the roots of an equation of the third degree

$$\Theta^3 - b_1\Theta^2 + b_2\Theta - b_3 = 0, \tag{5}$$

the b_i being the elementary symmetric functions of Θ_1, Θ_2, Θ_3:

$$\begin{aligned}b_1 &= \Theta_1 + \Theta_2 + \Theta_3 = 2\Sigma x_1 x_2 = 2p,\\ b_2 &= \Sigma \Theta_1\Theta_2 \quad\ = \Sigma x_1^2 x_2^2 + 3\Sigma x_1^2 x_2 x_3 + 6 x_1 x_2 x_3 x_4,\\ b_2 &= \Theta_1\Theta_2\Theta_3 \quad\ = \Sigma x_1^3 x_2^2 x_3 + 2\Sigma x_1^3 x_2 x_3 x_4\\ &\qquad + 2\Sigma x_1^2 x_2^2 x_3^2 + 4\Sigma x_1^2 x_2^2 x_3 x_4.\end{aligned}$$

b_2 and b_3 can be expressed in terms of the elementary symmetric functions σ_1, σ_2, σ_3, σ_4 of the x_i. We have (method of Section 26):

$$\begin{aligned}\sigma_2^2 &= \Sigma x_1^2 x_2^2 + 2\Sigma x_1^2 x_2 x_3 + 6 x_1 x_2 x_3 x_4 = p^2,\\ \sigma_1\sigma_3 &= \quad\ \Sigma x_1^2 x_2 x_3 + 4 x_1 x_2 x_3 x_4 = 0,\\ -4\sigma_4 &= \qquad\quad -4 x_1 x_2 x_3 x_4 = -4r\end{aligned}$$

$$b_2 = \Sigma x_1^2 x_2^2 + 3\Sigma x_1^2 x_2 x_3 + 6 x_1 x_2 x_3 x_4 = p^2 - 4r;$$

$$\begin{aligned}\sigma_1\sigma_2\sigma_3 &= \Sigma x_1^3 x_2^2 x_3 + 3\Sigma x_1^3 x_2 x_3 x_4 + 3\Sigma x_1^2 x_2^2 x_3^2 + 8\Sigma x_1^2 x_2^2 x_3 x_4 = 0,\\ -\sigma_1^2\sigma_4 &= \quad -\Sigma x_1^3 x_2 x_3 x_4 \qquad\qquad -2\Sigma x_1^2 x_2^2 x_3 x_4 = 0,\\ -\sigma_3^2 &= \qquad\qquad -\Sigma x_1^2 x_2^2 x_3^2 \quad -2\Sigma x_1^2 x_2^2 x_3 x_4 = -q^2\end{aligned}$$

$$b_3 = \Sigma x_1^3 x_2^2 x_3 + 2\Sigma x_1^3 x_2 x_3 x_4 + 2\Sigma x_1^2 x_2^2 x_3^2 + 4\Sigma x_2^2 x_2^2 x_3 x_4 = -q^2$$

In this way equation (5) becomes:

$$\Theta^3 - 2p\Theta^2 + (p^2 - 4r)\Theta + q^2 = 0.$$

This equation is known as the *cubic resolvent* of the equation of the fourth degree; according to "Cardano," its roots Θ_1, Θ_2, Θ_3 can be expressed as radicals. Each single Θ permits a group of 8 permutations; but all three of them permit only $\mathfrak{B}_4$; hence

$$\mathsf{K}(\Theta_1, \Theta_2, \Theta_3) = \Delta_1.$$

The field A_2 arises from A_1 by the adjunction of an element which does not permit all four substitutions of $\mathfrak{B}_4$, but only the identity and (say) the substitution (1 2) (3 4). Such an element is $x_1 + x_2$. We have

$$(x_1 + x_2)(x_3 + x_4) = \Theta_1 \quad \text{and} \quad (x_1 + x_2) + (x_3 + x_4) = 0,$$

and so, say

$$x_1 + x_2 = \sqrt{-\Theta_1}; \qquad x_3 + x_4 = -\sqrt{-\Theta_1}.$$

Similarly, we have

$$x_1 + x_3 = \sqrt{-\Theta_2}; \qquad x_2 + x_4 = -\sqrt{-\Theta_2};$$
$$x_1 + x_4 = \sqrt{-\Theta_3}; \qquad x_2 + x_3 = -\sqrt{-\Theta_3}.$$

These three irrationalities are not independent since their product is rational:

$$\begin{aligned}\sqrt{-\Theta_1}\cdot\sqrt{-\Theta_2}\cdot\sqrt{-\Theta_3} &= (x_1 + x_2)(x_1 + x_3)(x_1 + x_4)\\ &= x_1^3 + x_1^2(x_2 + x_3 + x_4) + x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4\\ &= x_1^2(x_1 + x_2 + x_3 + x_4) + \Sigma\, x_1x_2x_3\\ &= \Sigma\, x_1x_2x_3\\ &= -q.\end{aligned}$$

We need exactly two quadratic irrationalities in order to descend from $\mathfrak{B}_4$ to $\mathfrak{E}$, or to ascend from A to Σ; for $\mathfrak{B}_4$ is of order 4 and has a subgroup of order 2. As a matter of fact, the x_i may be determined rationally by the three elements Θ (which depend on two of them); for, evidently, we have

$$\begin{cases} 2x_1 = \sqrt{-\Theta_1} + \sqrt{-\Theta_2} + \sqrt{-\Theta_3},\\ 2x_2 = \sqrt{-\Theta_1} - \sqrt{-\Theta_2} - \sqrt{-\Theta_3},\\ 2x_3 = -\sqrt{-\Theta_1} + \sqrt{-\Theta_2} - \sqrt{-\Theta_3},\\ 2x_4 = -\sqrt{-\Theta_1} - \sqrt{-\Theta_2} + \sqrt{-\Theta_3}.\end{cases}$$

These are the solution formulae of the general equation of the fourth degree. By virtue of their derivation they are also valid for any special quartic equation.

NOTE: Since

$$\begin{aligned}\Theta_1 - \Theta_2 &= -(x_1 - x_4)(x_2 - x_3),\\ \Theta_1 - \Theta_3 &= -(x_1 - x_3)(x_2 - x_4),\\ \Theta_2 - \Theta_3 &= -(x_1 - x_2)(x_3 - x_4),\end{aligned}$$

the discriminant of the cubic resolvent is equal to the discriminant of the original equation. Here we have a simple tool for computing the discriminant of the quartic, since we already know that of the cubic equations. We find:

$$D = 16p^4r - 4p^3q^2 - 128p^2r^2 + 144pq^2r - 27q^4 + 256r^3.$$

EXERCISES. 1. The group of the cubic resolvent of a definite equation of the fourth degree is the factor group of the group of the original equation with respect to its intersection with the fours-group $\mathfrak{B}_4$.

2. What is the group of the equation

$$x^4 + x^3 + x + 1 = 0\ ?$$

[Cf. Ex. 4. Section 50, and the above Ex. 1.]

59. CONSTRUCTIONS WITH RULER AND COMPASS

We shall consider the question: *What geometrical constructions are possible by means of ruler and compass?*

Let there be given some elementary geometric objects (points, straight lines, or circles). The problem is to construct others from them which satisfy certain conditions.

Suppose that, in addition to the given figures, a Cartesian coordinate system is given. Then all given figures can be represented by numbers (coordinates), and the same is true for the figures to be constructed. If it is possible to construct the latter numbers (as line segments), the problem is solved. Thus everything reduces to the construction of segments from given segments. Let $a, b, \ldots$ be the given segments, and let x be the desired segment.

For the present we can state a sufficient condition for the constructibility:

Whenever a solution x of the problem is real and can be found by rational operations and (not necessarily real) square roots from the given segments $a, b, \ldots$, the segment x can be constructed with ruler and compass.

This theorem can be proved most conveniently as follows: We represent all complex numbers $p + iq$, which enter into the calculation of x, by points in a plane with rectangular coordinates p, q in the well-known manner,[9] and replace all operations to be performed by geometric constructions in this plane. The way in which this is done is generally known: Addition is vector addition, and subtraction is the inverse operation. In multiplication the angles are added, and the radii are multiplied; thus, if φ_1, φ_2 are the angles or arguments, and r_1, r_2 the radii or absolute values of the numbers to be multiplied, we have to construct the angle φ and the radius r for the product by means of the equations

$$\varphi = \varphi_1 + \varphi_2 \text{ and } r = r_1 r_2 \text{ or } 1 : r_1 = r_2 : r.$$

Division is the inverse operation again. Finally, in order to find the square root of a number with the absolute value r and the argument φ, we construct r_1, φ_1 from

$$\varphi = 2\varphi_1 \text{ or } \varphi_1 = \tfrac{1}{2}\varphi,$$

and

$$r = r_1^2 \text{ or } 1 : r_1 = r_1 : r.$$

[9] For the present we assume that the reader is familiar with the complex numbers. Later on in Chapter IX, we shall deal with the definition of real and complex numbers and discuss their rôle in abstract algebra.

Thus everything is reduced to well-known constructions with ruler and compass.[10]

The converse of the theorems just proved holds as well:

If a segment x can be constructed from given segments $a, b, \ldots$ by means of ruler and compass, then x can be expressed in terms of $a, b, \ldots$ by rational operations and square roots.

In order to prove this, let us examine more closely the operations which may be used in the construction. These are: Assumption of an arbitrary point (inside a given region); construction of a straight line through two points, of a circle with given center and radius, finally, of the intersection of two straight lines, of a straight line and a circle, or of two circles.

With the aid of our coordinate system all these operations can be followed algebraically. If we can assume an arbitrary point within a region, we may, in particular, assume that its coordinates are rational numbers. All other constructions lead to rational operations, except the two last ones (intersection of circles with straight lines or circles) which lead to quadratic equations, and hence to square roots. This proves the theorem.

We have to bear in mind that in a geometrical problem it is not essential whether we can find a construction for any *special* choice of the given points, but that a *general* construction is postulated which—with certain limitations—always yields a solution. Algebraically, this amounts to the fact that one and the same formula (which may contain square roots) always yields, within certain limitations, a meaningful solution x which satisfies the equations of the geometric problem. We may also say that the equations that determine x, and the square roots, etc., by which we solve the equations must remain meaningful when the given elements $a, b, \ldots$ are replaced by *indeterminates*. Consider, for example, the question whether the trisection of an angle can be performed by means of ruler and compass. This problem can, by virtue of the relation

$$\cos 3\varphi = 4\cos^3\varphi - 3\cos\varphi$$

be reduced to the solution of the equation

$$4x^3 - 3x = \alpha \qquad (\alpha = \cos 3\varphi). \tag{1}$$

Now, the question is not whether, for any special value of α, a solution of equation (1) by means of square roots can be found, but we ask whether a general solution formula of the equation exists, i.e., a solution formula that remains meaningful for an indeterminate α.

By the preceding discussion, we have reduced the geometrical problem of the possibility of a construction with ruler and compass to the following algebraic problem: What quantities x can be expressed in terms of given quantities $a, b \ldots$ by rational operations and square roots?

---

[10] We obtain another proof when we split all numbers involved into a real and imaginary part, and reduce, by Section 69, the complex square roots to real ones, which can be constructed in a well-known manner.

This question is easy to answer. Let $\mathfrak{R}$ be the field of rational functions of the given quantities $a, b, \ldots$. If x is to be expressed in terms of $a, b \ldots$ by rational operations and square roo's, then x must belong to a field which arises from $\mathfrak{R}$ by the successive adjunction of a finite number of square roots, i.e., to a field obtained by a finite number of extensions of degree 2. If, after adjoining each square root, we also adjoin the square roots of the conjugate field elements, then all extensions remain quadratic, and we get a normal extension field of degree 2^m in which x lies. Hence:

In order that the line segment x be constructible with ruler and compass, it is necessary that the real number x belongs to a normal extension field (of $\mathfrak{R}$) of degree 2^m.

This condition is also sufficient. For the Galois group of a field of degree 2^m is a group of order 2^m and, therefore, like any group of prime power degree, a *soluble* group (Section 46, Ex. 5). Hence there exists a composition series in which all composition factors are of order 2, and to this series corresponds, by the Fundamental Theorem of the Galois Theory, a chain of fields in which each field is of degree 2 over its predecessor. But an extension of degree 2 can always be obtained by the adjunction of a square root. Therefore, the quantity x can be expressed in terms of square roots, whence the theorem follows.

We proceed to apply these general theorems to some classical problems.

The problem of the *duplication of the cube* leads to the cubic equation

$$x^3 = 2,$$

which, by Eisenstein's criterion, is irreducible so that every root gives rise to an extension field of degree 3, which cannot be a subfield of a field of degree 2^m. *Therefore, the duplication of the cube is impossible with ruler and compass alone.*

The problem of trisecting an angle leads, as we saw before, to the equation

$$4x^3 - 3x - \alpha = 0,$$

where α is an indeterminate. It is easy to show the irreducibility of this equation in the rational domain of α: If the left member had a rational factor in α, it would also have an integral rational factor in α; but, evidently, a linear polynomial in α, whose coefficients have no common divisor is irreducible. From this we infer that an angle cannot be trisected by means of ruler and compass.

A more convenient algebraic form for the trisection equation is obtained by adjoining the element

$$i \sin 3\varphi = \sqrt{-(1 - \cos^2 3\varphi)}$$

to the rational domain of $\alpha = \cos 3\varphi$, and by finding the equation for

$$y = \cos\varphi + i \sin\varphi.$$

This equation is given by

$$(\cos\varphi + i \sin\varphi)^3 = \cos 3\varphi + i \sin 3\varphi,$$

or briefly by

$$y^3 = \beta.$$

It is easily seen directly from the geometric interpretation of the complex numbers that the trisection of the angle 3φ can be reduced to this pure equation.

The quantities x and y can be expressed in terms of one another by aid of square roots.

The squaring (quadrature) of the circle leads to the construction of the number π. Its impossibility will be proved if it can be shown that π does not satisfy any algebraic equation, in other words, that it is transcendental; for then π cannot lie in a finite extension field of the field of rationals. Regarding this proof, which does not belong to the domain of algebra, the reader is referred to G. Hessenberg's book: *Transzendenz von e und* π.

The construction of the regular polygons with a given circumscribed circle leads in case of h sides to the construction of

$$2\cos\frac{2\pi}{h} = \zeta + \zeta^{-1},$$

where ζ is the primitive h-th root of unity $e^{\frac{2\pi i}{h}}$. Since this sum is carried into itself only by the substitutions $\zeta \to \zeta$ and $\zeta \to \zeta^{-1}$ of the Galois group of the cyclotomic field, it generates a real subfield of degree $\frac{\varphi(h)}{2}$. The condition for its constructibility is that $\frac{\varphi(h)}{2}$, and so $\varphi(h)$, be a power of 2. Now for $h = 2^{\nu} q_1^{\nu_1} \dots q_r^{\nu_r}$ (where the q_i are odd primes) we have

$$\varphi(h) = 2^{\nu-1} q_1^{\nu_1-1} \cdots q_r^{\nu_r-1}(q_1-1)\cdots(q_r-1). \tag{2}$$

(In case $\nu = 0$ the first factor $2^{\nu-1}$ is missing.) Thus the condition is that only the first powers of the odd prime factors may divide h ($\nu_i = 1$) and, furthermore, that for every odd prime q_i dividing h the number $q_i - 1$ be a power to the base 2, i.e., every q_i must be of the form

$$q_i = 2^k + 1.$$

Which are the prime numbers of this form?

k cannot be divisible by an odd number $\mu > 2$; for from

$$k = \mu\nu, \qquad \mu \not\equiv 0 \pmod 2, \qquad \mu > 2$$

would follow that $(2^{\nu})^{\mu} + 1$ would be divisible by $2^{\nu} + 1$ and, therefore, would not be prime.

Thus we must have $k = 2^{\lambda}$ and

$$q_i = 2^{2^{\lambda}} + 1.$$

As a matter of fact, the values $\lambda = 0, 1, 2, 3, 4$ yield prime numbers q_i, namely

$$3, 5, 17, 257, 65537.$$

For $\lambda = 5$ and some larger λ (how many is unknown) $2^{2^{\lambda}} + 1$ is no longer prime; for example, $2^{2^5} + 1$ has 641 as a divisor.

As soon as the number h contains, besides powers of 2, only primes of the sequence 3, 5, 17, . . . to at most the first power, the regular polygon with h sides will be constructible (Gauss). The example of a polygon of 17 sides was treated in Section 49. The constructions of the triangle, the quadrilateral, pentagon, hexagon, octagon, and decagon are known. The regular heptagon and nonagon ($h = 7$ and 9) are no longer constructible, since they lead to cubic subfields in cyclotomic fields of degree 6.

EXERCISE. Show that in the casus irreducibilis the cubic equation

$$x^3 + px + q = 0$$

can always be reduced to the trisection equation (1) by a substitution $x = \beta x'$, and derive a solution formula for this cubic equation by means of trigonometric functions.

60. METACYCLIC EQUATIONS OF PRIME DEGREE

To an irreducible equation of prime degree q belongs a transitive permutation group of "degree" q, i.e., a transitive group $\mathfrak{G}$ of permutations of q objects 1, 2, . . . q. We shall investigate these groups and their normal divisors and, in particular, find out which transitive groups of degree q are metacyclic.

In Section 49 we already observed that the subgroup of $\mathfrak{G}_1$, which leaves the number 1 fixed, has the index q whence it follows that the order of $\mathfrak{G}$ must be divisible by the degree q. This conclusion holds even if q is not a prime.

If $\mathfrak{H}$ is a normal divisor of $\mathfrak{G}$, there are two possibilities:

Either $\mathfrak{H}$ is transitive, in which case the order of $\mathfrak{H}$ is again divisible by q, or $\mathfrak{H}$ is intransitive. Now let $\{1, 2, \ldots, k\}$ be a transitivity set of $\mathfrak{H}$, and σ a substitution in $\mathfrak{G}$, which carries the digit 1 into another digit i not belonging to the transitivity set; in this case $\sigma\{1, 2, \ldots, k\}$ will be a transitivity set of $\sigma\mathfrak{H}\sigma^{-1}$. But since $\mathfrak{H}$ is a normal divisor, we have $\sigma\mathfrak{H}\sigma^{-1} = \mathfrak{H}$; therefore, $\sigma\{1, 2, \ldots, k\}$ is again a transitivity set of $\mathfrak{H}$ which, moreover, consists of precisely k digits and contains the digit i. Since i was an arbitrary digit, all transitivity sets consist of the same number, i.e., of k digits; hence k is a (proper) divisor of q.

If q is a prime, as was assumed at the outset, the only possibility is $k = 1$; in this case, however, $\mathfrak{H}$ leaves fixed all digits 1, 2, . . . q. Hence:

A normal divisor $\mathfrak{H}$ of a transitive permutation group of prime degree q is either transitive or equal to $\mathfrak{E}$.

We proceed to prove the following theorem:

Any transitive metacyclic group $\mathfrak{G}$ of prime degree q can be written as a group of linear substitutions modulo q of the suitably numbered objects 1, 2, . . . , q:

$$\tau(z) \equiv az + b \pmod{q} \qquad (a \not\equiv 0 \,(q);\ z = 1, 2, \ldots, q).$$

Moreover, all substitutions with $a = 1$:

$$\sigma(z) \equiv z + b \qquad (b = 1, \ldots, q)$$

occur in the group.

PROOF. The order of the group $\mathfrak{G}$ is divisible by q. If it is equal to q, the group is cyclic; for the order of an arbitrary element σ distinct from the identity is a divisor of q and, therefore, can only be equal to q, and so σ already generates the complete group. The generating permutation σ must consist of a single cycle containing all the digits 1, 2, . . . , q; for otherwise the group would not be transitive. Therefore, with a suitable numbering we have

$$\sigma = (1\ 2 \ldots q)$$

so that

$$\sigma(z) \equiv z + 1 \pmod{q},$$
$$\sigma^b(z) \equiv z + b \pmod{q} \qquad (b = 1, \ldots, q).$$

In this case the theorem is proved. So we can perform an induction on the order of the group $\mathfrak{G}$ and assume that this order is a composite number $q \cdot j$, and that the theorem is (for a fixed q) valid for all groups of lower order.

Because of the solubility of $\mathfrak{G}$ there exists a soluble normal divisor $\mathfrak{H}$ of prime index distinct from $\mathfrak{E}$. By the preceding theorem, this normal divisor $\mathfrak{H}$ is transitive and therefore, by the induction hypothesis, a group of linear substitutions modulo q, which includes the group of the substitutions $z \to z + b$.

It is readily seen, that all substitutions $z \to z + b$ with $b \not\equiv 0$ are cycles of q terms. They are the only ones in the group $\mathfrak{H}$; for any other substitution $z \to az + b$ leaves fixed one element z, which can be determined from

$$az + b \equiv z,$$
$$(a - 1)z \equiv -b.$$

Now, if σ is the substitution

$$\sigma(z) \equiv z + 1,$$

and τ an arbitrary substitution in $\mathfrak{G}$, then $\tau\sigma\tau^{-1}$ is again a cycle of q terms and contained in $\mathfrak{H}$, and so again of the form

$$\tau\sigma\tau^{-1}(z) \equiv z + a.$$

Now, let $\tau^{-1}(z) = \zeta$, so that $z = \tau(\zeta)$,

$$\tau\sigma(\zeta) \equiv \tau(\zeta) + a,$$
$$\tau(\zeta + 1) \equiv \tau(\zeta) + a.$$

This holds for all ζ. By induction on ν we find

$$\tau(\zeta + \nu) \equiv \tau(\zeta) + \nu a,$$

and, in particular, for $\zeta = 0$, if we put $\tau(0) = b$,

$$\tau(\nu) \equiv \nu a + b.$$

Hence τ is a linear substitution modulo q. Since all the substitutions $\sigma(z) \equiv z + b$ are already contained in $\mathfrak{H}$, they are contained in $\mathfrak{G}$ as well. This completes the proof of the theorem.

Conversely, every group $\mathfrak{G}$ of linear substitutions modulo q which contains all the substitutions

$$\sigma(z) \equiv z + b$$

is soluble.

PROOF. The substitutions σ just mentioned form a normal divisor $\mathfrak{N}$ in $\mathfrak{G}$; for, with σ, every $\tau\sigma\tau^{-1}$ is a cycle of q terms, or the identity. In every coset $\mathfrak{N}\tau$, where τ represents the substitution

$$\tau(z) \equiv az + b$$

we have the substitution $\sigma^{-1}\tau$:

$$\sigma^{-1}\tau(z) \equiv az.$$

Two cosets are most conveniently composed by simply composing the substitutions $\tau'(z) \equiv az$. This is done by multiplying the coefficients a (which, accordingly, form a multiplicative group modulo q). Hence the cosets modulo $\mathfrak{N}$ form an Abelian group: $\mathfrak{G}/\mathfrak{N}$ is Abelian. Since also $\mathfrak{N}$ is Abelian, $\mathfrak{G}$ is soluble.

DEDUCTIONS. A linear substitution

$$\sigma(z) \equiv az + b,$$

distinct from the identity, leaves fixed at most one element; for the congruence

$$(a - 1)z \equiv -b \pmod{q}$$

has at most one solution unless $a \equiv 1$ and $b \equiv 0$. This means that *the only subgroup that leaves fixed two digits i, k is the identity group.*

The considerations used in our proofs also serve to determine all normal divisors of the linear group $\mathfrak{G}$. We have seen that every normal divisor (except $\mathfrak{E}$) must include the normal divisor $\mathfrak{N}$, and that in every coset modulo $\mathfrak{N}$ there exists a substitution $\tau(z) \equiv az$. Thus, it is sufficient to determine all subgroups in the multiplicative group of the $a \pmod{q}$ involved.

Now, the group of *all* residue classes $\not\equiv 0 \pmod{q}$ is cyclic and so is every subgroup. If the given group of residues is of order j, then to every divisor of j belongs just one subgroup.

If the objects of the permutations are the roots of an equation, and if $\mathfrak{G}$ is the Galois group of the equation, then our group theorems can at once be transformed into theorems concerning fields. We obtain:

The group of an irreducible metacyclic equation of prime degree q over the field K may, if suitable subscripts are affixed to the roots, always be regarded as a group of linear substitutions of the subscripts mod q. The degree of the decomposition field $\mathsf{K}(\alpha_1, \ldots, \alpha_q)$ is $q \cdot j$, where j is a divisor of $q-1$. There exists a normal intermediate field of degree j in which all other normal intermediate fields are contained. To every divisor of j belongs just one normal intermediate field. The intermediate field $\mathsf{K}(\alpha_i, \alpha_k)$, generated by two roots α_i, α_k is necessarily identical with the entire field $\mathsf{K}(\alpha_1, \ldots, \alpha_q)$ already.

EXERCISES. 1. A metacyclic irreducible equation of prime degree $q \neq 2$ over a real number field K has either only one real root; or all its roots are real.

2. An irreducible quintic with precisely three real roots is not solvable by radicals.

3. Prove with the aid of 2. that the equation

$$x^5 - 4x + 2 = 0$$

is not solvable by radicals. For determining the number of real roots, theorems of Chapter IX, e.g. Weierstrass' theorem and Rolle's theorem, may be used.

61. ESTABLISHMENT OF THE GALOIS GROUP. EQUATIONS WITH A SYMMETRIC GROUP

A method for actually forming—at least theoretically—the Galois group of an equation $f(x) = 0$ relative to a field Δ is the following:

Let the roots of the equation be $\alpha_1, \ldots, \alpha_n$. By means of the indeterminates $u_1, \ldots, u_n$ form the expression

$$\vartheta = u_1\alpha_1 + \cdots + u_n\alpha_n,$$

perform on it all permutations s_u of the indeterminates u and form the product

$$F(z, u) = \prod_s (z - s_u\vartheta).$$

Evidently, this product is a symmetric function of the roots, and therefore, by Section 26, it can be expressed in terms of the coefficients of $f(x)$. Now, decompose $F(z, u)$ into irreducible factors in $\Delta[u, z]$:

$$F(z, u) = F_1(z, u)\, F_2(z, u) \ldots F_r(z, u).$$

The permutations s_u which carry any of the factors, say F_1, into itself, form a group $\mathfrak{g}$. We now assert that $\mathfrak{g}$ *is exactly the Galois group of the given equation.*

PROOF. After adjoining all roots, F and therefore F_1 are decomposed into linear factors $z - \Sigma u_\nu \alpha_\nu$ with the roots α_ν as coefficients in any sequential order. We now affix subscripts to the roots in such fashion that F_1 contains the factor $z - (u_1\alpha_1 + \cdots + u_n\alpha_n)$. By s_u we shall hereafter denote any permutation of the u, and by s_α the same permutation of the α. Then, obviously, the product $s_u s_\alpha$ leaves invariant the expression $\vartheta = u_1\alpha_1 + \cdots + u_n\alpha_n$, i.e., we have

$$s_u s_\alpha \vartheta = \vartheta$$
$$s_\alpha \vartheta = s_u^{-1}\vartheta.$$

If s_u belongs to the group $\mathfrak{g}$, i.e., if it leaves F_1 invariant, then s_u transforms every linear factor of F_1, especially the factor $z-\vartheta$, into a linear factor of F_1 again. If, conversely, a permutation s_u transforms the factor $z-\vartheta$ into another linear factor of F_1, it transforms F_1 into a polynomial which is irreducible in $\Delta[u, z]$, and which is a divisor of $F(z, u)$, and so it transforms F_1 into one of the polynomials F_j. This F_j has a linear factor in common with F_1. Therefore, the permutation necessarily transforms F_1 into itself, which means that s_u belongs to $\mathfrak{g}$. Thus $\mathfrak{g}$ consists of the permutations of the u which transform $z-\vartheta$ into a linear factor of F_1 again.

The permutations s_α of the Galois group of $f(x)$ are characterized by the property that they transform the quantity

$$\vartheta = u_1\alpha_1 + \cdots + u_n\alpha_n$$

into its conjugates. This means: s_α transforms ϑ into an element satisfying the same irreducible equation as ϑ, i.e. s_α carries the linear factor $z-\vartheta$ into another linear factor of F_1. Now, $s_\alpha\vartheta = s_u^{-1}\vartheta$; hence s_u^{-1} carries the linear factor $z-\vartheta$ again into a linear factor of F_1, i.e., s_u^{-1} and so s_u belong to $\mathfrak{g}$. The converse is also true. Thus the Galois group consists of exactly the same permutations as the group $\mathfrak{g}$, except that they are performed on the α instead of the u.

From a practical point of view, this method for determining the Galois group is not of so much interest. However, the following interesting fact can be derived from it:

Let $\mathfrak{R}$ be an integral domain with identity, and let the Unique Factorization Theorem be valid for it. Let $\mathfrak{p}$ be a prime ideal in $\mathfrak{R}$, and let $\overline{\mathfrak{R}} = \mathfrak{R}/\mathfrak{p}$ be the residue class ring. Let the quotient field of $\mathfrak{R}$ and $\overline{\mathfrak{R}}$ be Δ and $\overline{\Delta}$. Let $f(x) = x^n + \cdots$ be a polynomial in $\mathfrak{R}[x]$, and let $\bar{f}(x)$ be the polynomial associated with it in the homomorphism $\mathfrak{R} \to \overline{\mathfrak{R}}$, assuming that neither has a double root. Then the Galois group $\bar{\mathfrak{g}}$ of the equation $\bar{f} = 0$ relative to $\overline{\Delta}$ (as a permutation group of the suitably arranged roots) is a subgroup of the Galois group $\mathfrak{g}$ of $f = 0$.

PROOF. By Section 23,

$$F(z, u) = \prod_s (z - s_u\vartheta)$$

can be decomposed in $\mathfrak{R}[z, u]$ into irreducible factors $F_1F_2\ldots F_k$ in $\Delta[z, u]$ with integral rational coefficients, and by virtue of the homomorphism this decomposition applies to $\overline{\mathfrak{R}}[z, u]$ as well so that

$$\overline{F}(z, u) = \overline{F}_1\overline{F}_2\ldots\overline{F}_k.$$

It might be that the factors $\overline{F}_1, \ldots$ are further decomposable. The permutations of $\mathfrak{g}$ carry F_1 and so $\overline{F}_1$ into themselves, the other permutations of the u carry $\overline{F}_1$ into $\overline{F}_2, \ldots, \overline{F}_k$. The permutations of $\mathfrak{g}$ carry an irreducible factor of $\overline{F}_1$ into itself so that they cannot carry $\overline{F}_1$ into $\overline{F}_2, \ldots, \overline{F}_k$, but must carry $\overline{F}_1$ into $\overline{F}_1$, which means that $\bar{\mathfrak{g}}$ is a subgroup of $\mathfrak{g}$.

The theorem is frequently used for determining the group $\mathfrak{g}$. In particular, we often choose the ideal $\mathfrak{p}$ in such manner that the polynomial $f(x)$ is decomposed mod $\mathfrak{p}$, since in this way the Galois group $\bar{\mathfrak{g}}$ of $\bar{f}$ can be determined more easily. Let, for example, $\mathfrak{R}$ be the ring of integers, and let $\mathfrak{p} = (p)$, p being a prime number. Let $f(x)$ be decomposed modulo p thus:

$$f(x) \equiv \varphi_1(x)\,\varphi_2(x) \ldots \varphi_h(x) \quad (p).$$

It follows that

$$\bar{f} = \bar{\varphi}_1 \bar{\varphi}_2 \cdots \bar{\varphi}_h.$$

The Galois group $\bar{\mathfrak{g}}$ of $\bar{f}(x)$ is always cyclic, since the automorphism group of a Galois field is always cyclic (Section 37). Let the generating permutation s of $\bar{\mathfrak{g}}$, written as a product of cycles, be

$$(1\ 2 \ldots j)\,(j+1 \ldots) \ldots$$

Since the transitivity sets of the group $\bar{\mathfrak{g}}$ correspond exactly to the irreducible factors of $\bar{f}$, the numbers occurring in the cycles $(1\ 2 \ldots j)$, $(\ldots)$, $\ldots$ must exactly denote the roots of $\bar{\varphi}_1$ and $\bar{\varphi}_2, \ldots$. Thus, as soon as the degrees $j, k, \ldots$ of $\varphi_1, \varphi_2, \ldots$ are known, the type of the substitution s is known as well: s consists of a cycle of j terms, of a cycle of k terms, etc. Since, with a suitable arrangement of the roots, $\bar{\mathfrak{g}}$ is a subgroup of $\mathfrak{g}$ by the above theorem, $\mathfrak{g}$ *must contain a permutation of the same type.*

Thus, for example, if a quintic with integral coefficients resolves modulo any prime number into an irreducible factor of the second and into one of the third degree, the Galois group contains a permutation of the type $(1\ 2)\ (3\ 4\ 5)$.

EXAMPLE. Consider the equation

$$x^5 - x - 1 = 0.$$

The left member is decomposable modulo 2 into

$$(x^2 + x + 1)\,(x^3 + x^2 + 1)$$

and is irreducible modulo 3; for if it had a linear or quadratic factor, it would have a factor in common with $x^9 - x$ (Section 37, Ex. 6), and would therefore have to have a factor in common with either $x^5 - x$ or $x^5 + x$, which evidently is not the case. Hence the group contains a cycle of five symbols and a product $(i\ k)$ $(l\ m\ n)$. The third power of the latter permutation is $(i\ k)$. This permutation, when transformed by $(1\ 2\ 3\ 4\ 5)$ and its powers, yields a chain of transpositions $(i\ k)$, $(k\ p)$, $(p\ q)$, $(q\ r)$, $(r\ i)$ which together generate the symmetric group. Hence the group $\mathfrak{g}$ is the *symmetric group.*

Now, we can prove the following theorem, which enables us to construct equations of any given degree with symmetric group: *A transitive permutation group of n objects containing a cycle of two symbols and a $(n-1)$-cycle, is the symmetric group.*

PROOF. Let $(1\ 2 \ldots n-1)$ be the $(n-1)$-cycle. By virtue of the transitivity, the cycle $(i\ j)$ can be transformed into $(k\ n)$, where k is one of the digits between 1 and $(n-1)$. The transformation of $(k\ n)$ by $(1\ 2 \ldots n-1)$ and its

powers yields all cycles $(1\ n)$, $(2\ n)$, . . . , $(n-1\ n)$, and these cycles together generate the symmetric group.

In order to construct, by this theorem, an equation of the n-th degree $(n > 3)$ with symmetric group, we first choose a polynomial f_1 of degree n, irreducible modulo 2, then a polynomial f_2 which resolves modulo 3 into an irreducible factor of degree $(n-1)$ and a linear factor, and finally a polynomial f_3 of degree n which resolves modulo 5 into a quadratic factor and into one or two factors of odd degree (all irreducible mod 5). All this is possible, since there exist polynomials of any degree which are irreducible modulo any prime number (Section 37, Ex. 6). Finally we choose f such that

$$f \equiv f_1 \pmod 2$$
$$f \equiv f_2 \pmod 3$$
$$f \equiv f_3 \pmod 5,$$

which is always possible. For example, it suffices to choose

$$f = -15f_1 + 10f_2 + 6f_3.$$

In this case the Galois group is transitive (since the polynomial is irreducible mod 2), contains a cycle of the type $(1\ 2\ \ldots\ n-1)$, and a cycle of order 2 multiplied by cycles of odd order. By raising this product to a suitable odd power, we obtain a pure cycle of order two, and we conclude, by the above theorem, that the Galois group is the symmetric group.

The method of construction just demonstrated is by far not the only one. We may, for instance, use Eisenstein's theorem (Section 24) in order to ensure the irreducibility of the equation and thus the transitivity of the group. When we deal with equations of odd degree $n > 3$, we may proceed in an even simpler manner by seeing to it that the equation resolves modulo 2 into factors of the degrees $(n-1)$ and 1, but modulo 3 into factors of the degrees $(n-2)$ and 2. In this case, the irreducibility is secured automatically. For all even degrees > 6 the same result can be achieved by decomposing modulo 3, as before, but modulo 5 into factors of the degrees 2, 3 and $n-5$. Other criteria and methods for forming equations of the kind desired may be found in an article by Ph. Furtwängler in *Math. Ann. Vol. 85, pp. 34-40.* It is in general an unsolved problem whether there are equations with rational coefficients having for their group an arbitrarily given permutation group (Cf. E. Noether: "Gleichungen mit vorgeschriebener Gruppe." *Math. Ann. Vol. 78, p. 221.*)

EXERCISES. 1. What is the group (with respect to the field of rationals) of the equation

$$x^4 + 2x^2 + x + 3 = 0?$$

2. Construct an equation of the sixth degree whose group is the symmetric group.

CHAPTER VIII

INFINITE FIELD EXTENSIONS

Every field arises from its prime field by a finite or infinite field extension. In Chapters V and VII we studied finite field extensions; in this chapter we shall enter into the study of infinite field extensions. We shall first deal with algebraic, and next with transcendental extensions.

All fields under consideration will be commutative.

62. ALGEBRAICALLY CLOSED FIELDS

It is natural that the *maximal* algebraic extensions, i.e., those which cannot be algebraically extended any more, play an important part among the algebraic extensions of a given field. In this section we shall prove the existence of such extensions.

For Ω to be such a maximal algebraic extension field it is necessary that every polynomial in $\Omega[x]$ resolve completely into linear factors; (for, otherwise, by Section 32, the field Ω could be extended by the adjunction of a root of a non-linear prime polynomial.) This condition is also sufficient. For if every polynomial can be decomposed into linear factors in $\Omega[x]$, then, if Ω' is an algebraic extension field, every element of Ω' must satisfy an equation in Ω ; hence (since the left member decomposes into linear factors) it must satisfy even a linear equation in Ω and, therefore, it must lie in Ω already. Hence we have $\Omega' = \Omega$, and Ω is maximal.

Therefore, we lay down the following definition:

A field is called algebraically closed if in $\Omega[x]$ *every polynomial resolves into linear factors.*

An equivalent definition is the following: Ω *is algebraically closed if every non-constant polynomial in* $\Omega[x]$ *possesses at least one root in* Ω, *or a linear factor in* $\Omega[x]$.

For if this condition is satisfied, and if we decompose an arbitrary polynomial $f(x)$ into prime factors, the latter can only be linear.

The "Fundamental Theorem of Algebra" to which we shall recur in Section 69, states that the field of complex numbers is algebraically closed. Another example of an algebraically closed field is the field of all complex algebraic numbers, i.e., of all those complex numbers which satisfy an equation with rational coefficients; for

the complex roots of an equation with algebraic coefficients are not only algebraic with respect to the field of algebraic numbers, but also algebraic with respect to the field of rationals and, therefore, are algebraic numbers themselves.

E. Steinitz proved that every field P can be extended to an algebraically closed field Ω. In his proof he makes use of the "well-ordering" (Wohlordnung) of the field P, which is basically a transcendental device. For this proof, which presupposes a good knowledge of the theory of sets, we refer the reader to Steinitz's original text quoted in the introduction of this book. We shall treat only the case of a countable field P. This is the most important case in algebra, and it is accessible by algebraic methods. Moreover, this special case already exhibits all essential algebraic features of Steinitz's proof.

We proceed to prove an existence theorem and a uniqueness theorem for algebraically closed fields.

EXISTENCE THEOREM. *Every countable field* P *possesses an algebraically closed algebraic extension field* Ω.

The proof of this theorem rests on a few lemmas:

LEMMA 1. *Let* Ω *be an algebraic extension field of* P. *A sufficient condition that* Ω *be algebraically closed is that all polynomials in* $\mathsf{P}[x]$ *resolve in* $\Omega[x]$ *into linear factors.*

PROOF. Let $f(x)$ be a polynomial in $\Omega[x]$. If $f(x)$ did not resolve into linear factors, we could adjoin a root α, and so obtain a larger field Ω'. α is algebraic with respect to Ω, and Ω is algebraic with respect to P; so α is algebraic with respect to P. Therefore, α is a root of a polynomial $g(x)$ in $\mathsf{P}[x]$. This polynomial, however, resolves in $\Omega[x]$ into linear factors. Therefore, α is a root of a linear factor in $\Omega[x]$, and therefore it lies in Ω, contrary to our hypothesis.

LEMMA 2. *If* P *is countable, so is the polynomial domain* $\mathsf{P}[x]$ *as well as every simple algebraic extension* $\mathsf{P}(\vartheta)$, *and a definite rule can be given for the counting as soon as the counting of* P *and the primitive element* ϑ *are known.*

PROOF. It will suffice to carry out the proof for the polynomial domain $\mathsf{P}[x]$; for, by Section 32, the elements of $\mathsf{P}(\vartheta)$ can be uniquely represented as polynomials of degree $<n$ in ϑ.

As the first element in our counting we take the zero; then, in lexicographical order, we count the finitely many polynomials of degree $\leqq 1$ with coefficients having numbers $\leqq 1$ in the counting of P; next, we count the finitely many polynomials of degree $\leqq 2$ with coefficients having numbers $\leqq 2$; next, those of degree $\leqq 3$ with coefficients having numbers $\leqq 3$, etc.

LEMMA 3. *If* $f(x)$ *is a polynomial of degree* n *over the countable field* P *and if* n *symbols* $\alpha_1, \ldots, \alpha_n$ *are given, then, according to a unique rule, a decomposition field* $\mathsf{P}(\alpha_1, \ldots, \alpha_n)$, *in which* $f(x)$ *is completely decomposed into linear factors* $(x - \alpha_\nu)$ *can be constructed and counted.*

PROOF. The construction of the decomposition field was already treated in Section 35. There the decomposition field was obtained by successive simple adjunctions of a root α_{k+1} to $\mathsf{P}(\alpha_1, \ldots, \alpha_k)$. Each step is a simple extension, and the counting is given by Lemma 2 each time as soon as the symbol α_{k+1} and the defining equation for α_{k+1} are known. Thus, the entire construction is uniquely determined, provided we specify that each time α_{k+1} shall be a root of that factor of $\dfrac{f(x)}{(x-\alpha_1)\cdots(x-\alpha_k)}$ irreducible in $\mathsf{P}(\alpha_1, \ldots, \alpha_k)$ which comes first in the counting of the polynomial domain $\mathsf{P}(\alpha_1, \ldots, \alpha_k)[x]$.

LEMMA 4. *If in an ordered sequence of fields every preceding field is a subfield of its successor, then the union of these fields is itself a field.*

PROOF. For any two elements α, β of the union there are two fields Σ_α, Σ_β which contain α and β, respectively, and one of these includes the other. In this comprehending field $\alpha+\beta$ and $\alpha\cdot\beta$ are defined, and these definitions coincide for all fields of the sequence which include α and β, since one of two such fields is always a subfield of the other. If we wish to prove, for example, the associative law

$$\alpha\beta\cdot\gamma=\alpha\cdot\beta\gamma,$$

we select from among the fields Σ_α, Σ_β, Σ_γ the largest one; α, β, and γ are contained in this field, and the associative law holds in it. In like manner all rules of operation may be proved.

PROOF OF THE EXISTENCE THEOREM. From Lemma 1 we see that, in order to construct an algebraically closed extension field Ω of P, we need only construct a field algebraic over P in which all polynomials of $\mathsf{P}[x]$ are completely decomposed.

Let the non-constant polynomials in $\mathsf{P}[x]$ be numbered according to Lemma 2. Let there be associated with every polynomial $f_\nu(x)$ of degree n, n new symbols $\alpha_{\nu 1}, \ldots, \alpha_{\nu n}$. We construct a countable decomposition field P_1 to $f_1(x)$, according to Lemma 3. Over P_1 we construct a decomposition field P_2 for f_2, etc. By Lemma 4, the union of all fields $\mathsf{P}_1, \mathsf{P}_2, \ldots$ is a field Ω. Since all fields P_ν are algebraic over P, so is Ω. In Ω all polynomials $f_1(x), f_2(x), \ldots$ resolve into linear factors. Thus, by Lemma 1, Ω is algebraically closed.

UNIQUENESS THEOREM. *Any two algebraically closed algebraic extension fields Ω, Ω' of a countable field P are equivalent.*

PROOF. Every element of Ω or Ω' is a root of a polynomial $f_\nu(x)$ in $\mathsf{P}[x]$, and every $f_\nu(x)$ has but a finite number of roots; therefore, both Ω and Ω' are countable. For we may first count all the roots of $f_1(x)$, then those of $f_2(x)$, etc., and always omit the roots that occurred before.[1]

[1] Since the counting of the roots of a polynomial $f(x)$ is not unique, we must use the axiom of choice at this point.

Let $\omega_1, \omega_2, \ldots$ be all the elements of Ω. We wish to extend, step by step, the identical automorphism of P to an isomorphism $\mathsf{P}(\omega_1, \ldots, \omega_n) \cong \mathsf{P}(\omega_1^*, \ldots, \omega_n^*)$, where $\omega_\nu^* \in \Omega'$. Suppose the isomorphism $\mathsf{P}(\omega_1, \ldots, \omega_{n-1}) \cong \mathsf{P}(\omega_1^*, \ldots, \omega_{n-1}^*)$ has already been constructed. ω_n is a root of an irreducible polynomial $f(x)$ over $\mathsf{P}(\omega_1, \ldots, \omega_{n-1})$. To this polynomial corresponds $f^*(x)$ over $\mathsf{P}(\omega_1^* \ldots, \omega_{n-1}^*)$ in the isomorphism. Let ω_n^* be the first root of $f^*(x)$ in the counting of Ω'. Then, by Section 35, the isomorphism

$$\mathsf{P}(\omega_1, \ldots, \omega_{n-1}) \cong \mathsf{P}(\omega_1^*, \ldots, \omega_{n-1}^*)$$

may be uniquely extended to an isomorphism

$$\mathsf{P}(\omega_1, \ldots, \omega_n) \cong \mathsf{P}(\omega_1^*, \ldots, \omega_n^*),$$

which carries ω_n into ω_n^*.

The extending sequence of isomorphisms constructed in this manner associates a certain element ω_n^* of Ω' with every element ω_n of Ω. To the sum $\omega_p + \omega_q$ and the product $\omega_p \cdot \omega_q$ correspond a sum and a product; for both ω_p and ω_q occur in a finite extension field $\mathsf{P}(\omega_1, \ldots, \omega_n)$ already. Hence Ω is isomorphic with a subfield Ω^* of Ω'. Since Ω is algebraiclly closed, Ω^* is also algebraically closed; therefore, every element of Ω' is already contained in Ω^*, i.e., we have $\Omega^* = \Omega'$ and $\Omega \cong \Omega'$.

The significance of the algebraically closed extension fields of a given field lies in the fact that they include all possible algebraic extensions apart from equivalent extensions, or more precisely:

If Ω is an algebraically closed algebraic extension field of P and Σ any algebraic extension field of P, there exists an extension field Σ_0 equivalent to Σ within Ω.

PROOF. Extend Σ to an algebraically closed algebraic extension field Ω'. The latter is algebraic with respect to P as well, and so equivalent to Ω. In a 1-isomorphism which carries Ω' into Ω and leaves all elements of P fixed, Σ goes into an equivalent subfield Σ_0 of Ω.

If we take the rational number field Γ as the underlying field P, the construction given in the proof of the main theorem yields a field Ω in denumerably many actually possible steps. This field is called *the field of all algebraic numbers* (cf. Section 69). Its subfields, i.e., the algebraic extensions of Γ, are called *algebraic number fields*. Proceding from the field $GF(p)$ of the residues modulo p, we construct in like manner a field $\Omega(p)$ which includes all Galois fields of characteristic p.

EXERCISE. Prove the existence and uniqueness of an extension field of P which arises by the adjunction of all zeros of a given (countable) set of polynomials in $\mathsf{P}[x]$.

63. SIMPLE TRANSCENDENTAL EXTENSIONS

As we know, every simple transcendental extension of a (commutative) field Δ is equivalent to the quotient field $\Delta(x)$ of the polynomial domain $\Delta[x]$. We may therefore restrict ourselves to the study of this quotient field

$$\Omega = \Delta(x).$$

The elements of Ω are rational functions

$$\eta = \frac{f(x)}{g(x)},$$

which can be assumed to be in lowest terms (f and g are relatively prime). The highest of the two degrees of $f(x)$ and $g(x)$ is called the *degree* of the function η.

THEOREM. *Every non-constant η of degree n is transcendental relative to Δ, and $\Delta(x)$ is algebraic of degree n relative to $\Delta(\eta)$.*

PROOF. Let the representation $\eta = \frac{f(x)}{g(x)}$ be in lowest terms. Then x satisfies the equation

$$g(x) \cdot \eta - f(x) = 0$$

with coefficients in $\Delta(\eta)$. These coefficients cannot all be zero. For if all of them were zero, and if a_k were a non-vanishing coefficient in $g(x)$, and b_k the coefficient of the same power of x in $f(x)$, we would have

$$a_k \eta - b_k = 0$$

so that $\eta = \frac{b_k}{a_k} =$ constant, contrary to hypothesis. Therefore, x is algebraic with respect to $\Delta(\eta)$.

If η were algebraic with respect to Δ, then x would be algebraic with respect to Δ, which is not the case. Hence η is transcendental.

x is a root of the polynomial in $\Delta(\eta)[z]$,

$$g(z)\eta - f(z)$$

of degree n. This polynomial is irreducible in $\Delta(\eta)[z]$; for if it were not, it would have to be reducible in $\Delta[\eta, z]$, according to Section 23; since it is linear in η, a factor would have to be independent of η and depend solely on z; but no such factor exists since $g(z)$ and $f(z)$ are relatively prime.

Hence x is algebraic of degree n with respect to $\Delta(\eta)$, whence the theorem $(\Delta(x):\Delta(\eta)) = n$ follows.

We note for future reference that the polynomial

$$g(z)\eta - f(z)$$

has no factor (in $\Delta[z]$) depending on z alone. This fact remains true when we replace η by its value $\frac{f(x)}{g(x)}$ and multiply by its denominator $g(x)$; thus, the polynomial in $\Delta[x, z]$,

$$g(z)f(x) - f(z)g(x)$$

has no factor depending on z alone.

The theorem just proved gives rise to three *deductions.*

1. The degree of a function $\eta=\frac{f(x)}{g(x)}$ depends only on the fields $\Delta(\eta)$ and $\Delta(x)$, but not on the particular choice of the generator x of the latter field.

2. $\Delta(\eta)=\Delta(x)$ holds only if η is of degree 1, i.e., if η is a linear fractional function. In other words: *All linear fractional functions of x, and only these functions, are field generators.*

3. An automorphism of $\Delta(x)$ which leaves fixed the elements of Δ must carry x again into a field generator. If, conversely, we carry x into another field generator $\bar{x}=\frac{ax+b}{cx+d}$, and every $\varphi(x)$ into $\varphi(\bar{x})$, then an automorphism arises which leaves fixed the elements of Δ. Therefore:

The automorphisms of $\Delta(x)$ relative to Δ are the fractional linear substitutions

$$\bar{x}=\frac{ax+b}{cx+d}, \qquad ad-bc\neq 0.$$

The following theorem is important for certain geometric investigations:

LÜROTH'S THEOREM: *Every intermediate field Σ with $\Delta\subset\Sigma\subseteq\Delta(x)$ is a simple transcendental extension: $\Sigma=\Delta(\vartheta)$.*

PROOF. The element x must be algebraic with respect to Σ; for if η is any element of Σ not lying in Δ, then x, as was shown before, is algebraic with respect to $\Delta(\eta)$, and, therefore, even more with respect to Σ. Let the polynomial irreducible in the polynomial domain $\Sigma[z]$ with the leading coefficient 1 and the root x be

$$f_0(z)=z^n+a_1z^{n-1}+\cdots+a_n. \tag{1}$$

We wish to determine the structure of this $f_0(z)$.

The a_i are rational functions of x. Multiplication by the l.c.m. of the denominators makes them integral rational functions and, at the same time, gives us a primitive polynomial with respect to x (cf. Section 23):

$$f(x,z)=b_0(x)z^n+b_1(x)z^{n-1}+\cdots+b_n(x).$$

Let the degree of this irreducible polynomial in x be m, and let the degree in z be n.

Not all the coefficients $a_i=\frac{b_i}{b_0}$ of (1) can be independent of x; for, otherwise, x would be algebraic with respect to Δ so that one of the coefficients, say

$$\vartheta=a_i=\frac{b_i(x)}{b_0(x)},$$

or, written in lowest terms,

$$\vartheta=\frac{g(x)}{h(x)}$$

must depend on x. The degrees of $g(x)$ and $h(x)$ are $\leqq m$. The (non-vanishing) polynomial

$$g(z) - \vartheta h(z) = g(z) - \frac{g(x)}{h(x)} h(z)$$

has the root $z = x$, and is therefore divisible by $f_0(z)$ in $\Sigma[z]$. If, by Section 23, we pass from these polynomials rational in x to integral polynomials primitive in x, this divisibility is preserved and we have

$$h(x)g(z) - g(x)h(z) = q(x, z)f(x, z).$$

In x the left-hand member is of degree $\leqq m$. On the right, however, f is of degree m already; hence it follows that the degree on the left is exactly m, and that $q(x, z)$ does not depend on x. But the left-hand side does not have a factor that depends on z alone (see above); hence $q(x, z)$ is a constant:

$$h(x)g(z) - g(x)h(z) = q \cdot f(x, z).$$

Thus, since the constant q does not matter, the form of $f(x, z)$ is determined. The degree of $f(x)$ in x is m. Thus, (for reasons of symmetry) the degree in z is also m so that $m = n$. At least one of the degrees of $g(x)$ and $h(x)$ must actually reach the maximum value m; therefore, ϑ, too, as a function of x, is precisely of degree m.

Thus, we have

$$(\Delta(x) : \Delta(\vartheta)) = m,$$

and, on the other hand,

$$(\Delta(x) : \Sigma) = m;$$

therefore, since Σ includes $\Delta(\vartheta)$,

$$(\Sigma : \Delta(\vartheta)) = 1,$$
$$\Sigma = \Delta(\vartheta)$$ Q.E.D.

The significance of Lüroth's theorem in geometry is as follows:

A plane (irreducible) algebraic curve $F(\xi, \eta) = 0$ is called *rational* if its points, except a finite number of them, can be represented in terms of rational parametric equations:

$$\xi = f(t),$$
$$\eta = g(t).$$

It may happen that every point of the curve (perhaps with a finite number of exceptions) belongs to several values of t. (Example: If we put

$$\xi = t^2,$$
$$\eta = t^2 + 1,$$

the same point belongs to t and $-t$). But by means of Lüroth's theorem this can always be avoided by a suitable choice of the parameter. For let Δ be a field containing the coefficients of the functions f, g, and let t, for the present, be an indeter-

minate. $\Sigma = \Delta(f, g)$ is a subfield of $\Delta(t)$. If t' is a primitive element of Σ, we have, for example,

$$f(t) = f_1(t') \quad \text{(rational)}$$
$$g(t) = g_1(t') \quad \text{(rational)}$$
$$t' = \varphi(f, g) = \varphi(\xi, \eta),$$

and we can verify easily that the new parametrization

$$\xi = f_1(t'),$$
$$\eta = g_1(t')$$

represents the same curve, while the denominator of the function $\varphi(x, y)$ vanishes only at a finite number of points of the curve so that to all points of the curve (apart from a finite number of them) there belongs only *one* t'- value.

EXERCISE. If the field $\Delta(x)$ is normal with respect to the subfield $\Delta(\eta)$, the polynomal (1) is decomposed in it into linear factors. All these linear factors arise, by fractional linear transformations of x, from one of them, say $z - x$. These linear transformations are characterized by the fact that they form a finite group and leave the function $\vartheta = \frac{g(x)}{h(x)}$ invariant.

64. THE DEGREE OF TRANSCENDENCE

Let Ω be an extension field of a fixed field P. An element v of Ω is called *algebraically dependent* on $u_1, \ldots, u_n$ if v is algebraic with respect to the field $\mathsf{P}(u_1, \ldots, u_n)$, i.e., if v satisfies an algebraic equation

$$a_0(u)v^g + a_1(u)v^{g-1} + \cdots + a_g(u) = 0$$

in which the coefficients $a_0(u), \ldots, a_g(u)$ are polynomials in $u_1, \ldots, u_n$ with coefficients in P, and if not all of them are zero.

The algebraic dependence relation has the following fundamental properties which are completely analogous to the fundamental properties of linear dependence (cf. Section 33):

FUNDAMENTAL THEOREM 1. *Every* u_i $(i = 1, \ldots, n)$ *is algebraically dependent on* $u_1, \ldots, u_n$.

FUNDAMENTAL THEOREM 2. *If* v *is algebraically dependent on* $u_1, \ldots, u_n$, *but not on* $u_1, \ldots, u_{n-1}$, *then* u_n *is algebraically dependent on* $u_1, \ldots, u_{n-1}, v$.

PROOF. Let us adjoin $u_1, \ldots, u_{n-1}$ to the underlying field. Then v is algebraically dependent on u_n, and therefore the following algebraic relation is valid:

(1) $$a_0(u_n)v^g + a_1(u_n)v^{g-1} + \cdots + a_g(u_n) = 0.$$

Arranging this equation according to powers of u_n, we have:

(2) $$b_0(v)u_n^h + b_1(v)u_n^{h-1} + \cdots + b_g(v) = 0.$$

By hypothesis, v is transcendental with respect to the underlying field $\mathsf{P}(u_1, \ldots, u_{n-1})$. Thus the polynomials $b_0(v), \ldots, b_g(v)$ are either identically zero in v or $\neq 0$. But not all of them can be identically zero in v, since otherwise the left member of (1) would also be identically zero in v, i.e., we would have $a_0(u_n) = a_1(u_n) = \cdots$

$= a_g(u_n) = 0$, which contradicts the hypothesis. Hence not all coefficients $b_h(v)$ in (2) are equal to zero; thus, by (2), u_n is algebraically dependent on v with respect to the underlying field $\mathsf{P}(u_1, \ldots, u_{n-1})$.

FUNDAMENTAL THEOREM 3. *If w is algebraically dependent on $v_1, \ldots, v_s$, and if every $v_j (j = 1, \ldots, s)$ is algebraically dependent on $u_1, \ldots, u_n$, then w is algebraically dependent on $u_1, \ldots, u_n$.*

PROOF. If w is algebraic over the field $\mathsf{P}(v_1, \ldots, v_s)$ and therefore over the field $\mathsf{P}(u_1, \ldots, u_n, v_1, \ldots, v_s)$, and if this field is itself algebraic over $\mathsf{P}(u_1, \ldots, u_n)$, then, by Section 35, w is also algebraic over $\mathsf{P}(u_1, \ldots, u_n)$. Q.E.D.

Now that it has been proved that the fundamental theorems of linear dependence are fulfilled, all corollaries set forth in Section 33 hold as well. We shall call the elements $u_1, \ldots, u_n$ *algebraically independent* if none of them depends algebraically on the others. Two systems $\{u_1, \ldots, u_n\}$ and $\{v_1, \ldots, v_s\}$ are called (algebraically) *equivalent* if every v_k is algebraically dependent on $u_1, \ldots, u_n$, and every u_i on $v_1, \ldots, v_s$. Every finite system $\{u_1, \ldots, u_n\}$ is equivalent to an algebraically independent subsystem (Third Corollary). Two equivalent algebraically independent systems $\{u_1, \ldots, u_r\}$ and $\{v_1, \ldots, v_s\}$ have the same number of elements (Fifth Corollary).

A set $\mathfrak{M}$ (in particular a field Ω) is said to be of *finite degree of transcendence over* P if all elements of the set are algebraically dependent on a finite number from among them. Then there exists an algebraic *basis for* $\mathfrak{M}$, i.e., an algebraically independent subset $\{u_1, \ldots, u_r\}$, all elements of which depend on $\mathfrak{M}$ algebraically. The number r of the basis elements is independent of the choice of the basis and is called the *degree of transcendence of the set* $\mathfrak{M}$. The degree of transcendence is the maximum number of the algbraically independent elements of the set. A subset of $\mathfrak{M}$ is at most of the same degree of transcendence as the complete set, and an algbraic basis of the subset may be completed to an algebraic basis of the complete set (cf. Section 33, Ex. 2).

THEOREM. *The elements $u_1, \ldots, u_r$ are algebraically independent if, and only if, it follows necessarily from*

$$f(u_1, \ldots, u_r) = 0$$

where f is a polynomial with coefficients in P, that all coefficients of this polynomial vanish.

PROOF. Since $f(u_1, \ldots, u_r) = 0$ entails the identical vanishing of the polynomial f, it is clear that no u_i can be algebraically dependent on the other u_j. Let, conversely, $u_1, \ldots, u_r$ now be algebraically independent. If

$$f(u_1, \ldots, u_r) = 0 ,$$

and if we arrange the polynomial f according to powers of u_r, it follows that the coefficients $f_i(u_1, \ldots, u_{r-1})$ of this polynomial are equal to zero. If these coefficients are arranged according to powers of u_{r-1}, then, by similar conclusions, it finally follows that all coefficients of the polynomial f must be equal to zero.

If $u_1, \ldots, u_r$ are algebraically independent, then, by this theorem, they are not mutually connected by any algebraic equations whatsoever. Therefore, they are also called *independent transcendentals.* Thus the degree of transcendence of a field Ω with respect to P is the maximum number of the independent transcendentals contained in Ω.

If $u_1, \ldots, u_r$ are algebraically independent, and if $z_1, \ldots, z_r$ are indeterminates over P, then every polynomial $f(z_1, \ldots, z_r)$ with coefficients in P can be placed into a one-to-one correspondence with a polynomial $f(u_1, \ldots, u_r)$. Hence $\mathsf{P}[z_1, \ldots, z_r] \cong \mathsf{P}[u_1, \ldots, u_r]$. From the isomorphism of the polynomial rings follows the isomorphism of their quotient fields:

$$\mathsf{P}(z_1, \ldots, z_r) \cong \mathsf{P}(u_1, \ldots, u_r).$$

Accordingly, all algebraic properties of the independent transcendentals $u_1, \ldots, u_r$ are the same as those of r indeterminates $z_1, \ldots, z_r$.

A field extension which arises by the adjunction of a finite number or of an infinite number of algebraically independent elements $u_1, \ldots, u_r$ is called a *pure transcendental extension.*

If Ω is an extension field of finite degree of transcendence over P, and if $u_1, \ldots, u_r$ is an algebraic basis of Ω, then every element of Ω is algebraic over $\mathsf{P}(u_1, \ldots, u_r)$, that is:

Every extension field Ω of finite degree of transcendence over P can be obtained by a pure transcendental extension $\mathsf{P}(u_1, \ldots, u_r)$ and a subsequent algebraic extension.

Steinitz, with the aid of his well-ordering methods, proved this theorem not only for extensions of finite degree of transcendence, but for arbitrary transcendental extension fields. This theorem by Steinitz gives us a clear idea of the structure of transcendental extension fields. However, the fields of finite degree of transcendence, to which we have limited our investigation, are by far the most important ones for the applications.

EXERCISE. An extension obtained by two successive extensions of the finite degrees of transcendence s and t is of degree of transcendence $s + t$.

65. DIFFERENTIATION OF ALGEBRAIC FUNCTIONS

The definition of the derivative of a polynomial $f(x)$, as laid down in Section 20, may directly be applied to rational functions of one indeterminate

$$\varphi(x) = \frac{f(x)}{g(x)}$$

with coefficients in a field P. For if we form

$$\varphi(x+h) - \varphi(x) = \frac{f(x+h)\,g(x) - f(x)\,g(x+h)}{g(x)\,g(x+h)},$$

the numerator of this fraction becomes zero for $h=0$; therefore, it contains the factor h. Dividing both sides by h, we obtain

$$\frac{\varphi(x+h)-\varphi(x)}{h}=\frac{q(x,h)}{g(x)\,g(x+h)}. \tag{1}$$

The right-hand side is a rational function of h, which has a certain value for $h=0$ since the denominator does not vanish. This value is called the *differential quotient* or the *derivative* $\varphi'(x)$ of the rational function $\varphi(x)$:

$$\varphi'(x)=\frac{d\varphi(x)}{dx}=\frac{q(x,0)}{g(x)^2}. \tag{2}$$

For actual computation of $q(x,0)$, we develop the numerator of the right-hand side of (1) according to ascending powers of h, divide by h, put $h=0$, and obtain the result

$$q(x,0)=f'(x)g(x)-f(x)g'(x),$$

which, when substituted in (2), yields the well-known formula for the differentiation of a quotient, viz.

$$\frac{d}{dx}\frac{f(x)}{g(x)}=\frac{f'(x)\,g(x)-f(x)\,g'(x)}{g(x)^2}.$$

Let $R(u_1,\ldots,u_n)$ be a rational function; let $R'_1,\ldots,R'_n$ be its partial derivatives with respect to the indeterminates $u_1,\ldots,u_n$, and let $\varphi_1,\ldots,\varphi_n$ be rational functions of x.

We shall now prove the *law of total differentiation*

$$\frac{d}{dx}R(\varphi_1,\ldots,\varphi_n)=\sum_1^n R'_\nu(\varphi_1,\ldots,\varphi_n)\frac{d\varphi_\nu}{dx}. \tag{3}$$

For this purpose we put, according to the definition of the derivative,

$$\varphi_\nu(x+h)-\varphi_\nu(x)=h\psi_\nu(x,h),\quad \psi_\nu(x,0)=\varphi'_\nu(x),$$

and

$$\begin{aligned}
&R(u_1+h_1,\ldots,u_n+h_n)-R(u_1,\ldots,u_n)\\
&=\sum_{\nu=1}^n\{R(u_1+h_1,\ldots,u_\nu+h_\nu,u_{\nu+1},\ldots,u_n)-\\
&\qquad\qquad -R(u_1+h_1,\ldots,u_\nu,u_{\nu+1},\ldots,u_n)\}\\
&=\sum_{\nu=1}^n h_\nu S_\nu(u_1+h_1,\ldots,u_\nu,h_\nu,u_{\nu+1},\ldots,u_n),
\end{aligned} \tag{4}$$

where

$$S_\nu(u_1,\ldots,u_\nu,0,u_{\nu+1},\ldots,u_n)=R'_\nu(u_1,\ldots,u_n).$$

If we substitute

$$u_\nu=\varphi_\nu(x),\quad h_\nu=\varphi_\nu(x+h)-\varphi_\nu(x)=h\psi_\nu(x,h)$$

in the identity (4) and divide by h, it follows that

$$\left\{\begin{array}{l}\dfrac{R(\varphi_1(x+h),\ldots,\varphi_n(x+h))-R(\varphi_1(x),\ldots,\varphi_n(x))}{h}\\ =\sum_{\nu=1}^{n}\psi_\nu(x,h)\,S_\nu(\varphi_1+h\psi_1,\ldots,\varphi_\nu,h\psi_\nu,\varphi_{\nu+1},\ldots,\varphi_n).\end{array}\right.$$

If we put $h=0$ on the right-hand side, it follows that

$$\frac{d}{dx}R(\varphi_1,\ldots,\varphi_n)=\sum\varphi_\nu'(x)R_\nu'(\varphi_1,\ldots,\varphi_n),$$

which proves (3).

We shall now attempt to extend the theory of differentiation to algebraic functions of a variable x. By an *algebraic function of the indeterminate* x we mean an arbitrary element η of an algebraic extension field of $\mathsf{P}(x)$. We merely make the assumption that η is separable with respect to $\mathsf{P}(x)$.

Thus, let the algebraic function η be a root of a separable polynomial $F(x,y)$ irreducible over $\mathsf{P}(x)$:

$$F(x,\eta)=0.$$

Let the derivatives of $F(x,y)$ with respect to x and y be denoted by F_x' and F_y'. Due to the separability, $F_y'(x,y)$ has no root in common with $F(x,y)$; thus we have

$$F_y'(x,\eta)\neq 0.$$

We expect from a reasonable definition of the derivative $d\eta/dx$ that the law of total differentiation holds for the polynomial $F(x,y)$ so that

$$F_x'(x,\eta)+\frac{d\eta}{dx}F_y'(x,\eta)=0.$$

Therefore we *define*

$$\frac{d\eta}{dx}=-\frac{F_x'(x,\eta)}{F_y'(x,\eta)}. \tag{5}$$

We see at once that the definition is independent of the choice of the defining polynomial $F(x,y)$; for if we replace $F(x,y)$ by $F(x,y)\cdot\psi(x)$, where $\psi(x)$ is any rational function of x, then $F_x'(x,\eta)$ and $F_y'(x,\eta)$ in (5) are replaced by

$$F_x'(x,\eta)\cdot\psi(x)+F(x,\eta)\cdot\psi'(x)=F_x'(x,\eta)\cdot\psi(x)$$

and

$$F_y'(x,\eta)\cdot\psi(x)$$

which implies that the quotient (5) remains unaltered.

If, in particular, $\eta=c$ is a constant in P, x does not occur in the defining equation of η at all so that $\dfrac{dc}{dx}=0$.

Let ζ be an element of the field $\mathsf{P}(x,\eta)$, i.e., a rational function of x and η and a polynomial in η:

$$\zeta=\varphi(x,\eta).$$

We wish to prove the law of total differentiation for this function φ :

$$\frac{d\zeta}{dx}=\varphi_x'(x,\eta)+\varphi_y'(x,\eta)\frac{d\eta}{dx}, \tag{6}$$

where φ'_x and φ'_y are the derivatives of $\varphi(x, y)$ with respect to x and y. I or this purpose we form the defining equation of ζ:

$$G(x, \zeta) = 0.$$

We may assume $G(x, \zeta)$ to be a polynomial in x and ζ. We substitute in this polynomial the expression $\varphi(x, \eta)$ for ζ, and replace η by the indeterminate y. The arising polynomial in y has η as a root and is therefore divisible by $F(x, y)$:

$$G(x, \varphi(x, y)) = Q(x, y)F(x, y).$$

Partial differentiation of this identity with respect to x and y according to the law of total differentiation (3) yields:

$$\begin{cases} G'_x(x, \varphi(x, y)) + G'_z(x, \varphi(x, y))\varphi'_x(x, y) = QF'_x + Q'_x F(x, y) \\ G'_z(x, \varphi(x, y))\varphi'_y(x, y) = QF'_y + Q'_y F(x, y). \end{cases}$$

Now we replace y by η again, thus making the terms with $F(x, y)$ vanish and put, according to definition (3),

$$F'_x(x, \eta) = -F'_y(x, \eta) \cdot \frac{d\eta}{dx}$$

$$G'_x(x, \zeta) = -G'_z(x, \zeta) \cdot \frac{d\zeta}{dx}.$$

Thus we obtain

$$\begin{cases} -G'_z(x, \zeta) \cdot \frac{d\zeta}{dx} + G'_z(x, \zeta)\varphi'_x(x, \eta) = -Q(x, \eta)F'_y(x, \eta) \cdot \frac{d\eta}{dx} \\ G'_z(x, \zeta)\varphi'_y(x, \eta) = Q(x, \eta)F'_y(x, \eta). \end{cases}$$

When we multiply the second equation by $\frac{d\eta}{dx}$, add it to the first, and divide the sum by G'_z, it follows that

$$-\frac{d\zeta}{dx} + \varphi'_x(x, \eta) + \varphi'_y(x, \eta) \cdot \frac{d\eta}{dx} = 0,$$

which proves (6).

Now that the special case (6) has been taken care of, the proof of the general *law of total differentiation* is no longer *difficult. If* $\eta_1, \ldots, \eta_n$ *are separable algebraic functions of* x *in a field, and if* $R(u_1, \ldots, u_n)$ *is a polynomial with the derivatives* R'_ν, *then*

$$\frac{d}{dx} R(\eta_1, \ldots, \eta_n) = \sum_1^n R'_\nu(\eta_1, \ldots, \eta_n) \frac{d\eta_\nu}{dx}. \tag{7}$$

PROOF. Let ϑ be a primitive element of the separable extension field $\mathsf{P}(x, \eta_1, \ldots, \eta_n)$ of $\mathsf{P}(x)$. Then all η_ν are rationally expressible in terms of x and ϑ:

$$\eta_\nu = \varphi_\nu(x, \vartheta).$$

If $\varphi'_{\nu x}$ and $\varphi'_{\nu t}$ are the derivatives of $\varphi_\nu(x, t)$ with respect to x and t, we have, by (6),

$$\frac{d\eta_\nu}{dx} = \varphi'_{\nu x}(x, \vartheta) + \varphi'_{\nu y}(x, \vartheta) \cdot \frac{d\vartheta}{dx};$$

similarly, if R'_x and R'_t are the derivatives of the function $R(\varphi_1(x,t),\ldots,\varphi_n(x,t))$, we have

$$\frac{d}{dx}R(\eta_1,\ldots,\eta_n)=\frac{d}{dx}R(\varphi_1(x,\vartheta),\ldots,\varphi_n(x,\vartheta))$$
$$=R'_x(x,\vartheta)+R'_t(x,\vartheta)\cdot\frac{d\vartheta}{dt}.$$

But by (3) we have

$$R'_x(x,t)=\sum_1^n R'_\nu(\varphi_1(x,t),\ldots,\varphi_n(x,t))\varphi'_{\nu x}(x,t)$$
$$R'_t(x,t)=\sum_1^n R'_\nu(\varphi_1(x,t),\ldots,\varphi_n(x,t))\varphi'_{\nu t}(x,t);$$

hence

$$\frac{d}{dx}R(\eta_1,\ldots,\eta_n)=\sum_1^n R'_\nu(\varphi_1(x,\vartheta),\ldots,\varphi_n(x,\vartheta))\left\{\varphi'_{\nu x}(x,\vartheta)+\varphi'_{\nu t}(x,\vartheta)\cdot\frac{d\vartheta}{dt}\right\}$$
$$=\sum_1^n R'_\nu(\eta_1,\ldots,\eta_n)\frac{d\eta_\nu}{dx}.$$

Important special cases of the general law (7) are:

(8) $$\frac{d}{dx}(\eta+\zeta)=\frac{d\eta}{dx}+\frac{d\zeta}{dx},$$

(9) $$\frac{d}{dx}\eta\zeta=\eta\frac{d\zeta}{dx}+\frac{d\eta}{dx}\zeta,$$

(10) $$\frac{d}{dx}\frac{\eta}{\zeta}=\frac{1}{\zeta^2}\left(\zeta\frac{d\eta}{dx}-\eta\frac{d\zeta}{dx}\right),$$

(11) $$\frac{d}{dx}\eta^r=r\eta^{r-1}\frac{d\eta}{dx}.$$

The definition (5) of the derivative is of course applicable not only when x is an indeterminate, but whenever x is a transcendental element with respect to the underlying field P, and when η is separable and algebraic over $\mathsf{P}(x)$. In this case we write ξ rather than x. Thus, in a field of degree of transcendence 1 over P all elements η, insofar as they are separable over $\mathsf{P}(\xi)$, can be differentiated with respect to the transcendental element ξ.

If η and ζ are algebraically dependent on ξ, the field $\mathsf{P}(\xi,\eta,\zeta)$ is of degree of transcendence 1 over P If η is transcendental over P, ζ is algebraically dependent on η; thus we can form $\frac{d\zeta}{d\eta}$. If

(12) $$G(\eta,\zeta)=0$$

is the defining equation of ζ over $\mathsf{P}(\eta)$, and if G'_y and G'_z are the partial derivatives of $G(y,z)$, then

(13) $$G'_y(\eta,\zeta)+G'_z(\eta,\zeta)\frac{d\zeta}{d\eta}=0.$$

If, on the other hand, we differentiate (12) with respect to ξ, then, by the law of total differentiation, we obtain

$$G'_y(\eta, \zeta)\frac{d\eta}{d\xi} + G'_z(\eta, \zeta)\frac{d\zeta}{d\xi} = 0. \tag{14}$$

When we multiply (13) by $\frac{d\eta}{d\xi}$ and subtract (14) from the product, we obtain the *chain rule*:

$$\frac{d\zeta}{d\xi} = \frac{d\zeta}{d\eta}\frac{d\eta}{d\xi}. \tag{15}$$

If, in particular, $\zeta = \xi$, then (15) yields

$$\frac{d\xi}{d\eta}\cdot\frac{d\eta}{d\xi} = 1. \tag{16}$$

Thus, we have derived all the rules of ordinary differential calculus for algebraic functions of one variable by purely algebraic methods without employing any limit concepts whatever.

CHAPTER IX

REAL FIELDS

Aside from the algebraic properties of the numbers in an algebraic number field, there are certain non-algebraic properties, such as absolute values $|a|$, reality, positiveness, which play a part in the theory of algebraic fields. That these properties cannot be defined uniquely with the aid of the algebraic operations $+$ and $\cdot$ is illustrated by the following example:

Let w be a real, and so iw a purely imaginary root of the equation $x^4 = 2$. In the isomorphism

$$\Gamma(w) \cong \Gamma(iw)$$

all algebraic properties are preserved; but this isomorphism carries the real number w into the purely imaginary iw, the positive number $w^2 = \sqrt{2}$ into the negative number $(iw)^2 = -\sqrt{2}$, while the number $1 + \sqrt{2}$ of absolute value > 1 goes into the number $1 - \sqrt{2}$ of absolute value < 1.

Nevertheless, in the course of our investigations we shall see that there is something algebraic in these non-algebraic properties. In the field of algebraic numbers (i.e., in the algebraically closed extension field belonging to Γ) we can characterize by algebraic properties not *one* subfield, but a whole family of subfields, each of which is algebraically equivalent to the field of real algebraic numbers. For a particular choice of such a field, whose elements may then be called "real," the absolute values and the positiveness can be defined algebraically. For any finite algebraic number field, the definitions of absolute values, reality, etc. become *finitely ambiguous*.

Before entering into the study of this algebraic theory, we shall discuss the (transcendental) introduction of real and complex numbers customary in analysis. The reason for this procedure is not so much the fact that it is a logical necessity to begin with it, but because the problems involved in the purely algebraic theory become clearer as soon as we know what real and complex numbers actually are, and because we can at the same time discuss the important basic concepts of ordering and of a fundamental sequence.

66. ORDERED FIELDS

The subject of this section is an axiomatic investigation of a first non-algebraic property, namely the "positiveness," and of the "ordering," which rests on the former.

A (commutative) field K *shall be called "ordered" if the property of positiveness* (>0) *is defined for its elements, and if it satisfies the following postulates:*

1. For every element a in K, just one of the relations

$$a=0, \quad a>0, \quad -a>0$$

is valid.

2. If $a>0$ and $b>0$, then $a+b>0$ and $ab>0$.

If $-a>0$, we say: a is *negative.*

The ordering relation $a>b$ in an ordered field is now defined by

$$a>b, \text{ in words: } a \text{ is greater than } b$$
$$(\text{or } b<a, \text{ in words: } b \text{ is less than } a)$$
$$\text{if } a-b>0.$$

We can readily show that the set-theoretical ordering axioms are fulfilled. For we have, for any two elements a, b, either $a<b$ or $a=b$ or $a>b$. From $a>b$ and $b>c$ follows $a-b>0$ and $b-c>0$ and so

$$a-c=(a-b)+(b-c)>0,$$

so that $a>c$. Furthermore, just as in Section 3, $a>b$ implies $a+c>b+c$, and if $c>0$, it also implies $ac>bc$. Finally, if a and b are positive, $a>b$ always implies $a^{-1}<b^{-1}$ (and vice versa), since

$$ab(b^{-1}-a^{-1})=a-b.$$

If the absolute value $|a|$ of an element a in an ordered field is defined as the non-negative among the elements a, $-a$, the following rules for absolute values hold:

$$|ab|=|a|\cdot|b|,$$
$$|a+b|\leqq|a|+|b|.$$

The first rule can be readily verified for the four possible cases, viz.

$$a\geqq 0, \quad b\geqq 0;$$
$$a\geqq 0, \quad b<0;$$
$$a<0, \quad b\geqq 0;$$
$$a<0, \quad b<0.$$

Evidently, for $a\geqq 0$, $b\geqq 0$ the second rule holds with the equality sign, since in this case both sides are equal to the non-negative number $a+b$, and similarly for $a<0$, $b<0$, in which case both sides are equal to the non-negative number $-(a+b)$. Hence only the second and the third of our four cases remain to be considered. It suffices to consider one of them, namely $a\geqq 0$, $b<0$. Here we have

$$a+b<a<a-b=|a|+|b|,$$
$$-a-b\leqq -b\leqq a-b=|a|+|b|,$$

and so

$$|a+b|\leqq|a|+|b|.$$

Furthermore, we have

$$a^2 = (-a)^2 = |a|^2 \geqq 0,$$

with the equality sign only for $a = 0$. From this follows that a sum of squares is always $\geqq 0$; it is equal to 0 only if all summands vanish individually.

In particular, the element $1 = 1^2$ is always positive, and so is every sum $n \cdot 1 = 1 + 1 + \cdots + 1$. Therefore, we cannot have $p \cdot 1 = 0$ if p is a prime. Hence: *The characteristic of an ordered field is zero.*

LEMMA. *If* K *is the quotient field of the ring* $\mathfrak{R}$, *and if* $\mathfrak{R}$ *is ordered, then there is one, and only one, way of ordering* K *so that the ordering of* $\mathfrak{R}$ *is preserved.*

For let K be ordered in the desired manner. An arbitrary element of K is of the form $a = \frac{b}{c}$ (b and c in $\mathfrak{R}$ and $c \neq 0$).

From $$\frac{b}{c} > 0, \text{ or } = 0, \text{ or } < 0$$

follows at once upon multiplication by c^2 that

$$bc > 0, \text{ or } = 0, \text{ or } < 0, \text{ respectively.}$$

Therefore, any possible ordering of K is uniquely determined by that of $\mathfrak{R}$. Conversely, it can be readily seen that the stipulation

$$\frac{b}{c} > 0, \text{ if } bc > 0$$

actually defines an ordering of K which preserves the ordering of $\mathfrak{R}$.

In particular, the field of rationals Γ can be ordered in only one way, since the ring C of integers, evidently, is capable of the natural ordering only. Thus we have $\frac{m}{n} > 0$, provided $m \cdot n$ is a natural number.

Two ordered fields are called *order-isomorphic* if there exists an isomorphism of the two fields which carries positive elements always into positive elements.

The ordering of a field is called *Archimedean* [1] if there exists a "natural number" $n > a$ for every field element a. In this case there exists also a number $-n < a$ for every a, and a fraction $\frac{1}{n} < a$ for every positive a. The ordering of the rational number field Γ is Archimedean. If the ordering of a field is not Archimedean, there exist "infinitely large" elements, larger than any rational number, and "infinitely small" elements which are smaller than any positive rational number but larger than zero.

Bibliography on non-Archimedean ordered fields:

E. Artin, and O. Schreier: "Algebraische Konstruktion reeller Körper." *Abh. Math. Sam. Hamburg Vol. 5 (1926) pp. 83-115.*

Baer, R.: "Über nichtarchimedisch geordnete Körper." *Sitzungsber. Heidelb. Ak., 8. Abhandlung, 1927.*

[1] The "Archimedean axiom" in geometry runs as follows: Starting from a given point P ("zero point") a given line segment PQ ("unity segment") can always be laid off in the direction PR a number of times so that the last end point lies beyond any given point R.

EXERCISES. 1. Let a polynomial $f(t)$ with rational coefficients be called positive if the coefficient of the highest power of the indeterminate t is positive. Show that an ordering of the polynomial ring $\Gamma[t]$ and, therefore, of the quotient field $\Gamma(t)$ is thus defined, and that the latter ordering is non-Archimedean (t is "infinitely large").

2. Let

$$f(x) = x^n + a_1 x^{n-1} + \cdots + a_n,$$

where the a_i are taken from an ordered field K. Let M be the largest of the elements 1 and $|a_1| + \cdots + a_n|$. Show that

$$f(s) > 0 \text{ for } s > \mathsf{M}$$
$$(-1)^n f(s) > 0 \text{ for } s < -\mathsf{M}.$$

Thus, if $f(x)$ has roots in K, they lie within the range $-\mathsf{M} \leqq s \leqq \mathsf{M}$.

3. With the designations given in Ex. 2 let $R > 0$, and let $-S$ be the sum of the negative quantities from among $a_1, \frac{a_2}{R}, \frac{a_3}{R^2}, \ldots, \frac{a_n}{R^{n-1}}$. Let M_R be the greater of the numbers R and S. Then, for $s > \mathsf{M}_R$ we always have $f(s) > 0$. (For $R = 1$ we obtain an improvement of the upper bound M of Ex. 2.)

4. Again let $f(x) = x^n + a_1 x^{n-1} + \cdots + a_n$, let all $a_\nu \geqq -c$, and $c \geqq 0$. Show that $f(s) > 0$ for $s \geqq 1 + c$.

[Use the inequality $s^m \geqq c(s^{m-1} + s^{m-2} + \cdots + 1)$.]

By replacing x by $-x$, determine in like manner a bound $-1 - c'$, so that

$$(-1)^n f(s) > 0 \text{ for } s < -1 - c'.$$

If, in addition to the leading coefficient 1, $a_1, \ldots, a_r$ are positive, the bound $1 + c$ may be replaced by $1 + \frac{c}{1 + a_1 + \cdots + a_r}$.

5. For $a > b > 0$ we have $a^n > b^n > 0$ (where n is a natural number). In any ordered field K the polynomial $x^n - c$ has at most one positive root $\sqrt[n]{c}$. If n is odd, it cannot have more than one root at all; if n is even, it has at most two roots w and $-w$. If both $\sqrt[n]{c}$ and $\sqrt[n]{d}$ exist, and if $0 < c < d$, then $\sqrt[n]{c} < \sqrt[n]{d}$.

67. DEFINITION OF THE REAL NUMBERS

Let K be an ordered field, and $\mathfrak{M}$ a non-empty set of elements in K. If all elements of $\mathfrak{M}$ are less than or equal to a fixed quantity s in K, s is called an *upper bound* of $\mathfrak{M}$, and $\mathfrak{M}$ is said to be *bounded from above*. If there exists an upper bound less than any other upper bound of $\mathfrak{M}$, the former is called the *least upper bound* of the set $\mathfrak{M}$.

If all elements of $\mathfrak{M}$ are greater than or equal to a fixed quantity s' in K, s' is called a *lower bound* of $\mathfrak{M}$, and $\mathfrak{M}$ is said to be *bounded from below*.

In the rational number field Γ not every set $\mathfrak{M}$ bounded from above has a least upper bound. EXAMPLE: Let $\mathfrak{M}$ be the set of positive numbers with squares less than 3. Any positive rational number with a square greater than 3 is an upper bound of $\mathfrak{M}$. A number of Γ with square 3 does not exist, since x^2-3 is irreducible in Γ. If r is a positive rational number, and if $r^2>3$, then

$$r'=\frac{r+3r^{-1}}{2}$$

is a smaller positive number with a square >3 ; for we have

$$r'=\frac{r+3r^{-1}}{2}>\frac{r}{2}>0,$$
$$r'=\frac{r+3r^{-1}}{2}<\frac{r+r^2r^{-1}}{2}=r,$$
$$r'^2=\left(\frac{r+3r^{-1}}{2}\right)^2=\left(\frac{r-3r^{-1}}{2}\right)^2+3\geqq 3, \text{ and so } >3.$$

Therefore, for every upper bound r in Γ there exists a smaller upper bound Consequently, no least upper bound exists.

Let us now try to find for every ordered field K an ordered extension field Ω , in which every non-empty set bounded from above has a least upper bound. If, in particular, K is the field of rational numbers, Ω will become the well-known field of "real numbers." Among the various constructions of the field Ω known from the foundation of analysis we choose Cantor's construction by "fundamental sequences."

An infinite sequence of elements $a_1, a_2, \ldots$ in an ordered field K is called a *fundamental sequence* $\{a_\nu\}$ if, for every positive element ε of K , there exists an integer $n=n(\varepsilon)$ such that

(1) $$|a_p-a_q|<\varepsilon \quad \text{for} \quad p>n,\ q>n.$$

For $q=n+1$ it follows from (1) that

$$|a_p|\leqq|a_q|+|a_p-a_q|<|a_{n+1}|+\varepsilon=\mathsf{M} \qquad \text{for} \quad p>n.$$

Hence every fundamental sequence is bounded from above and from below.

Sums and products of fundamental sequences are defined by

$$c_n=a_n+b_n; \quad d_n=a_nb_n.$$

We show that the sum and the product are themselves fundamental sequences. For every ε there exists an n_1 such that

$$|a_p-a_q|<\tfrac{1}{2}\varepsilon \qquad \text{for} \quad p>n_1,\ q>n_1$$

and an n_2 such that

$$|b_p-b_q|<\tfrac{1}{2}\varepsilon \qquad \text{for} \quad p>n_2,\ q>n_2.$$

If n is the largest of the numbers n_1 and n_2, it follows that

$$|(a_p+b_p)-(a_q+b_q)|<\varepsilon \qquad \text{for} \quad p>n,\ q>n.$$

Similarly, there exist an M_1 and an M_2 such that

$$|a_p|<\mathsf{M}_1 \qquad \text{for} \quad p>n_1,$$
$$|b_p|<\mathsf{M}_2 \qquad \text{for} \quad p>n_2,$$

and, furthermore, for every ε there exists an $n' \geqq n_2$ and an $n'' \geqq n_1$ such that

$$|a_p - a_q| < \frac{\varepsilon}{2M_2} \qquad \text{for} \quad p > n',\ q > n';$$

$$|b_p - b_q| < \frac{\varepsilon}{2M_1} \qquad \text{for} \quad p > n'',\ q > n''.$$

Hence it follows upon multiplication by $|a_q|$ and $|b_p|$, respectively, that

$$|a_p b_p - a_q b_p| < \frac{\varepsilon}{2} \qquad \text{for} \quad p > n',\ q > n',$$

$$|a_q b_p - a_q b_q| < \frac{\varepsilon}{2} \qquad \text{for} \quad p > n'',\ q > n'',$$

and, therefore, if n is the largest of the numbers n' and n'', we have

$$|a_p b_p - a_q b_q| < \varepsilon \qquad \text{for} \quad p > n,\ q > n.$$

Addition and multiplication of fundamental sequences evidently fulfill all postulates for a ring; hence: The *fundamental sequences form a ring* $\mathfrak{o}$.

A fundamental sequence $\{a_p\}$ which "converges to 0," i.e., in which, for every ε, there exists an n such that

$$|a_p| < \varepsilon \qquad \text{for} \quad p > n,$$

is called a *null sequence*. We proceed to show the following:

The null sequences form an ideal $\mathfrak{n}$ *in the ring* $\mathfrak{o}$.

PROOF. If $\{a_p\}$ and $\{b_p\}$ are null sequences, then, for every ε there exists an n_1 and an n_2 such that

$$|a_p| < \tfrac{1}{2}\varepsilon \qquad \text{for} \quad p > n_1,$$

$$|b_p| < \tfrac{1}{2}\varepsilon \qquad \text{for} \quad p > n_2;$$

if n is the largest of the numbers n_1 and n_2, this implies

$$|a_p - b_p| < \varepsilon \qquad \text{for} \quad p > n.$$

Hence $\{a_p - b_p\}$ is a null sequence as well. If, furthermore, $\{a_p\}$ is a null sequence, and $\{c_p\}$ any fundamental sequence, we determine an n' and an M such that

$$|c_p| < M \qquad \text{for} \quad p > n',$$

and for every ε an $n = n(\varepsilon) \geqq n'$, such that

$$|a_p| < \frac{\varepsilon}{M} \qquad \text{for} \quad p > n.$$

Then it follows that

$$|a_p c_p| < \varepsilon \qquad \text{for} \quad p > n;$$

so $\{a_p c_p\}$ is a null sequence.

Let the residue class ring $\mathfrak{o}/\mathfrak{n}$ be called Ω. We shall show that Ω *is a field*, i.e., that the congruence

$$a x \equiv 1 \,(\mathfrak{n}) \tag{2}$$

has a solution in $\mathfrak{o}$ for $a \not\equiv 0\,(\mathfrak{n})$. Here 1 is the identity of $\mathfrak{o}$, i.e., the fundamental sequence $\{1, 1, \ldots\}$.

An n and a $\eta > 0$ must exist such that

$$|a_q| \geqq \eta \qquad \text{for} \quad q > n.$$

for if, for all n and all $\eta > 0$, we would have

$$|a_q| < \eta \qquad (q > n),$$

then, for a given η, we could take n so large that for $p > n$, $q > n$ we would have

$$|a_p - a_q| < \eta;$$

hence

$$|a_p| < 2\eta$$

for all $p > n$; so the sequence $\{a_p\}$ would be a null sequence, contrary to the hypothesis.

The fundamental sequence $\{a_p\}$ remains in the same residue class modulo $\mathfrak{n}$, if we replace $a_1, \ldots, a_n$ by η. If we denote these n new elements η again by $a_1, \ldots, a_n$, we have for *all* p:

$$|a_p| \geqq \eta, \text{ in particular } a_p \neq 0.$$

Now $\{a_p^{-1}\}$ is a fundamental sequence. For there exists an n for every ε such that

$$|a_q - a_p| < \varepsilon\eta^2 \qquad \text{for } p > n,\ q > n.$$

Now, if we had $|a_p^{-1} - a_q^{-1}| \geqq \varepsilon$ for a $p > n$ and a $q > n$, then, upon multiplication by $|a_p| \geqq \eta$ and $|a_q| \geqq \eta$, it would follow that

$$|a_q - a_p| = |a_p a_q (a_p^{-1} - a_q^{-1})| \geqq \varepsilon\eta^2,$$

which is not the case. Therefore,

$$|a_p^{-1} - a_q^{-1}| < \varepsilon \qquad \text{for } p > n,\ q > n.$$

Obviously, the fundamental sequence $\{a_p^{-1}\}$ solves the congruence (2).

The field Ω contains, in particular, those residues mod $\mathfrak{n}$ which are represented by fundamental sequences of the form

$$\{a, a, a, \ldots\}.$$

They form a subring K' of Ω isomorphic with K; for to every a of K there corresponds such a residue class, to different a correspond different residue classes, and to the sum and the product correspond the sum and product, respectively. If we now identify the elements of K' with those of K, Ω becomes an extension field of K.

A fundamental sequence $\{a_p\}$ is called *positive* if there exists an $\varepsilon > 0$ in K, and an n such that

$$a_p > \varepsilon \qquad \text{for } p > n.$$

Clearly, sum and product of two fundamental sequences are themselves positive. Similarly, the sum of a positive sequence $\{a_p\}$ and a null sequence $\{b_p\}$ is always positive; this can be shown by choosing an n large enough so that

$$a_p > \varepsilon \qquad \text{for } p > n,$$
$$|b_p| < \tfrac{1}{2}\varepsilon \qquad \text{for } p > n$$

and by concluding that $a_p + b_p > \frac{1}{2}\varepsilon$ for $p > n$. Therefore, all sequences of a residue class mod $\mathfrak{n}$ are positive if one of them is positive. In this case the residue class itself is called *positive*. A residue class k is called negative, if $-k$ is positive.

If neither $\{a_p\}$ nor $\{-a_p\}$ is positive, then, for every $\varepsilon > 0$ and every n, there exists an $r > n$ and an $s > n$, such that

$$a_r \leqq \varepsilon \quad \text{and} \quad -a_s \leqq \varepsilon.$$

If we choose a sufficiently large n so that, for $p > n$, $q > n$, we have

$$|a_p - a_q| < \varepsilon,$$

then, by first taking $q = r$ and an arbitrary $p > n$, we conclude

$$a_p = (a_p - a_q) + a_r < \varepsilon + \varepsilon = 2\varepsilon,$$

and next, by taking $q = s$ and an arbitrary $p > n$,

$$-a_p = (a_q - a_p) - a_s < \varepsilon + \varepsilon = 2\varepsilon.$$

Consequently

$$|a_p| < 2\varepsilon \qquad \text{for } p > n.$$

Hence $\{a_p\}$ is a null sequence.

Therefore, either $\{a_p\}$ is positive, or $\{-a_p\}$ is positive, or $\{a_p\}$ is a null sequence. Consequently, every residue class mod $\mathfrak{n}$ is either positive, negative, or zero. Since the sum and the product of positive residue classes are themselves positive, we infer:

Ω is an ordered field.

We see at once that the ordering of K is preserved in Ω.

If a sequence $\{a_p\}$ defines an element α, and a sequence $\{b_p\}$ an element β of Ω, it always follows from

$$a_p \geqq b_p \qquad \text{for } p > n$$

that $\alpha \geqq \beta$. For if we had $\alpha < \beta$, and so $\beta - \alpha > 0$, then we would have an ε and an m for the fundamental sequence $\{b_p - a_p\}$ such that

$$b_p - a_p > \varepsilon > 0 \qquad \text{for } p > m.$$

If we choose $p = m + n$ here, we are led to a contradiction to the hypothesis $a_p \geqq b_p$. It is useful to remember that $a_p > b_p$ does not imply $\alpha > \beta$, but only $\alpha \geqq \beta$.

The fact that every fundamental sequence is bounded from above implies that, for every element ω of Ω, there exists a greater element s of K. If the ordering of K is *Archimedean*, there exists an integer $n > s$; hence for every ω there exists an $n > \omega$, i.e., *the ordering of Ω is Archimedean.*

In the field Ω itself we can again define the concepts of absolute value, fundamental sequence, and null sequence. The null sequences again form an ideal. If a sequence $\{\alpha_p\}$ is congruent to a constant sequence $\{\alpha\}$ modulo this ideal, i.e., if $\{\alpha_p - \alpha\}$ is a null sequence, we say: the sequence $\{\alpha_p\}$ converges to the limit α. In symbols:

$$\lim_{p \to \infty} \alpha_p = \alpha \quad \text{or briefly} \quad \lim \alpha_p = \alpha.$$

The fundamental sequences $\{a_p\}$ of K which were employed in the definition of the elements of Ω may of course be regarded as fundamental sequences in Ω; for K is contained in Ω. We proceed to prove: *If the sequence $\{a_p\}$ defines the*

element α *of* Ω, *then* $\lim a_p = \alpha$. To prove this we observe that for every positive ε in Ω there exists a smaller positive ε' in K, and for this ε' there again exists an n so that, for $p > n$, $q > n$ the relation

$$|a_p - a_q| < \varepsilon'$$

is always valid, i.e., that both $a_p - a_q$ and $a_q - a_p$ are smaller than ε'. If we now let p remain fixed, and q tend to ∞, it follows that $a_p - \alpha$ and $\alpha - a_p$ are both $\leqq \varepsilon'$ so that

$$|a_p - \alpha| \leqq \varepsilon' < \varepsilon.$$

Hence $\{a_p - \alpha\}$ is a null sequence.

We proceed to show that the field Ω cannot be extended any more by fundamental sequences, since every fundamental sequence $\{\alpha_p\}$ already has a limit in Ω. (Cauchy's Convergence Theorem).

In the proof we may assume that in the sequence $\{\alpha_p\}$ two succeeding elements α_p, α_{p+1} are always distinct from one another; for if this is not the case, we can either choose a subsequence consisting of the α_p, distinct from their α_{p-1} (the convergence of the subsequence, of course, implies at once the convergence of the given sequence), or the sequence, starting at a certain point, remains constant: $\alpha_p = \alpha$ for $p > n$. In the latter case it is obvious that $\lim \alpha_p = \alpha$.

We now put

$$|\alpha_p - \alpha_{p+1}| = \varepsilon_p.$$

Since $\{\alpha_p\}$ was a fundamental sequence, $\{\varepsilon_p\}$ is a null sequence.[2] By hypothesis $\varepsilon_p > 0$.

For every α_p we now choose an approximating a_p such that

$$|a_p - \alpha_p| < \varepsilon_p.$$

This is possible, since α_p itself was defined by a fundamental sequence $\{a_{p1}, a_{p2}, \ldots\}$ with the limit α_p. Furthermore, for every ε there exists an n' such that

$$|\alpha_p - \alpha_q| < \tfrac{1}{3}\varepsilon \qquad \text{for} \quad p > n',\ q > n',$$

and an n'' such that

$$\varepsilon_p < \tfrac{1}{3}\varepsilon \quad \text{for} \quad p > n''.$$

If n is the greater of the two numbers n' and n'', then, for $p > n$, $q > n$, all three absolute values $|a_p - \alpha_p|$, $|\alpha_p - \alpha_q|$ and $|\alpha_q - a_q|$ are less than $\frac{1}{3}\varepsilon$, so that

$$|a_p - a_q| \leqq |a_p - \alpha_p| + |\alpha_p - \alpha_q| + |\alpha_q - a_q| < \tfrac{1}{3}\varepsilon + \tfrac{1}{3}\varepsilon + \tfrac{1}{3}\varepsilon = \varepsilon.$$

Thus the a_p form a fundamental sequence in K which defines an element ω of Ω. The sequence $\{\alpha_p\}$ differs from this fundamental sequence only by a null sequence $\{a_p - \alpha_p\}$, and therefore has the same limit ω.

We proceed to prove the *Theorem on the Least Upper Bound* for the case of an

---

[2] The foregoing part of the proof merely served to secure the existence of a null sequence which will be needed hereafter. In the Archimedean case it would have been simpler to set $\varepsilon_p = 2^{-p}$; however, we wish to furnish a perfectly general proof. In the non-Archimedean case $\{2^{-p}\}$ is not a null sequence.

Archimedean ordering.

Every non-empty set $\mathfrak{M} \subset \Omega$ bounded from above has a least upper bound in Ω.

PROOF. Let s be an upper bound of $\mathfrak{M}$, and M an integer $> s$, furthermore, μ an arbitrary element of $\mathfrak{M}$, and m an integer $> -\mu$. Then

$$-m < \mu < M.$$

For every natural number p we now form the finite number of fractions $k \cdot 2^{-p}$ (where k is an integer) which lie "between" $-m$ and M:

$$-m \leqq k \cdot 2^{-p} \leqq M.$$

Let a_p be the smallest among those fractions which are, like M, upper bounds of the set $\mathfrak{M}$. Then $a_p - 2^{-p}$ is no longer an upper bound. Therefore, for every $q > p$ we have

$$a_p - 2^{-p} < a_q \leqq a_p. \tag{4}$$

Thence it follows that

$$|a_p - a_q| < 2^{-p},$$

so that

$$|a_p - a_q| < 2^{-n} \quad \text{for} \quad p > n,\ q > n. \tag{5}$$

For a given ε we can always find an integer $h > \varepsilon^{-1}$ and, moreover, a $2^n > h > \varepsilon^{-1}$. Then $2^{-n} < \varepsilon$. Thus, (5) implies that $\{a_p\}$ is a fundamental sequence. This sequence defines an element ω of Ω. Furthermore, it follows from (4) that

$$a_p - 2^{-p} \leqq \omega \leqq a_p.$$

ω is an upper bound of $\mathfrak{M}$, i.e., all elements μ of $\mathfrak{M}$ are $\leqq \omega$. For if we had $\mu > \omega$, we could find a number $2^p > (\mu - \omega)^{-1}$, and we would have $2^{-p} < \mu - \omega$. If we add $a_p - 2^{-p} \leqq \omega$, it follows that $a_p < \mu$, which is a contradiction, since a_p is an upper bound of $\mathfrak{M}$.

ω is the least upper bound of $\mathfrak{M}$. For if σ were a smaller upper bound, we could again find a number p such that $2^{-p} < \omega - \sigma$. Since $a_p - 2^{-p}$ is not an upper bound of $\mathfrak{M}$, there exists a μ in $\mathfrak{M}$ such that $a_p - 2^{-p} < \mu$.
This implies

$$a_p - 2^{-p} < \sigma.$$

Upon addition to the foregoing we obtain

$$a_p < \omega,$$

which is false. Therefore, ω is the least upper bound of $\mathfrak{M}$.

Thus, for every ordered field K the above construction affords a uniquely determined ordered extension field Ω, for which the theorem on the least upper bound is valid if K is Archimedean. If, in particular, K is the field of rational numbers, Ω is the field of *real numbers*. Thus, a real number, in this theory, is defined as a residue class modulo $\mathfrak{n}$ in the domain of fundamental sequences of

rational numbers.

EXERCISES. 1. Prove the following properties of the limit concept:

a) If $\{\alpha_n\}$ and $\{\beta_n\}$ are convergent sequences, then

$$\lim(\alpha_n \pm \beta_n) = \lim \alpha_n \pm \lim \beta_n,$$
$$\lim \alpha_n \beta_n = \lim \alpha_n \cdot \lim \beta_n.$$

b) If $\lim \beta_n \neq 0$, and all $\beta_n \neq 0$, then

$$\lim(\beta_n^{-1}) = (\lim \beta_n)^{-1}$$

c) A subsequence of a convergent sequence converges to the same limit.

2. The only way (except for equivalent extensions) to extend an Archimedean ordered field to another such field, in which Cauchy's convergence theorem is valid, is the above construction by means of fundamental sequences.

3. Every Archimedean ordered field is similarly-isomorphic to a subfield of the field of real numbers.

4. Every real number s can be represented by an infinite decimal

$$s = a_0 + \sum_{\nu=1}^{\infty} a_\nu 10^{-\nu} \left(\text{i.e., } s = \lim_{n \to \infty} (a_0 + \sum_{\nu=1}^{n} a_\nu 10^{-\nu}) \right) \quad (0 \leqq a_\nu < 10).$$

68. ZEROS OF REAL FUNCTIONS

Let P be the field of real numbers. We shall now consider real-valued functions $f(x)$ of a real variable x. Such a function is called *continuous* at $x = a$, if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|f(a+h) - f(a)| < \varepsilon \qquad \text{for } |h| < \delta.$$

It is easy to prove that the sums and the products of continuous functions are themselves continuous (cf. the corresponding proof for fundamental sequences in Section 67). Since the constants and the function $f(x) = x$ are continuous everywhere, all polynomials in x constitute continuous functions of x.

Weierstrass' Nullstellensatz for continuous functions reads as follows:

Let $f(x)$ be a function, continuous in the interval $a \leqq x \leqq b$; if $f(a) < 0$ and $f(b) > 0$, then the function has a zero between a and b.

PROOF. Let c be the least upper bound of all x between a and b, for which $f(x) < 0$. Then there are three possibilities:

1. $f(c) > 0$. In this case $c > a$, and there exists a $\delta > 0$ such that for $0 < h < \delta$ we have

$$f(c-h) - f(c) | < f(c),$$
$$f(c) - f(c-h) < f(c),$$

i.e.,

$$f(c-h) > 0,$$
$$f(x) > 0 \qquad \text{for } c - \delta < x \leqq c.$$

Therefore, $c - \delta$ is an upper bound for the x for which $f(x) < 0$. But c was the

least upper bound. Hence this case is impossible.

2. $f(c) < 0$. In this case $c < b$, and there exists a $\delta > 0$ such that for $0 < h < \delta$, e.g., for $h = \frac{1}{2}\delta$, we have

$$f(c+h) - f(c) < -f(c),$$
$$f(c+h) < 0.$$

Therefore, c is not an upper bound for all x for which $f(x) < 0$. Therefore, this case is impossible, too.

3. $f(c) = 0$ is the only case remaining. Hence $f(x)$ has c as a zero.

Weierstrass' Nullstellensatz for polynomials is the base of all theorems on real roots of algebraic equations. Later we shall apply it to fields other than the field of real numbers, namely, to the so-called "real closed fields." All theorems of this section rest exclusively on Weierstrass' Nullstellensatz for polynomials, and are, accordingly, valid for the more general fields to be discussed at a later stage.

DEDUCTIONS. 1. *For* $d > 0$ *the polynomial* $x^n - d$, *where n is a natural number, always has a positive root.*

For $x^n - d < 0$ for $x = 0$, and if x is large, $\left(\text{e.g., } x > 1 + \frac{d}{n}\right)$ we have $x^n - d > 0$.

Furthermore, from $a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + \cdots + b^{n-1})$ follows that, for $a > b > 0$, we also have $a^n > b^n$ so that there exists only one positive root of the equation $x^n = d$. This root is denoted by $\sqrt[n]{d}$. In the case $n = 2$ ("square root") we simply write $\sqrt{d}$. Furthermore we have $\sqrt[n]{0} = 0$. $a > b \geqq 0$ now implies $\sqrt[n]{a} > \sqrt[n]{b}$, for if we had $\sqrt[n]{a} \leqq \sqrt[n]{b}$ it would follow that $a \leqq b$.

2. *Every polynomial of odd degree has a root in* P.

For by Ex. 2, Section 66, there exists an M such that $f(M) > 0$ and $f(-M) < 0$.

We now turn to the *computation of the real roots of a polynomial* $f(x)$. By computation we mean, in accordance with the definition of real numbers, an arbitrarily close approximation by rational numbers.

We already saw in Section 66 (Ex. 2) how to inclose the real roots of $f(x)$ by bounds: If

$$f(x) = x^n + a_1 x^{n-1} + \cdots + a_n,$$

and if M is the largest of the numbers 1 and $|a_1| + \cdots + |a_n|$, then all roots lie between $-M$ and $+M$. [For $x > M$ the value of $f(x)$ is > 0 and has the same sign as $(-1)^n$ for $x < -M$.] M may be replaced by a (possibly larger) rational number, to be called M again, and then the interval $-M \leqq x \leqq M$ can be divided into arbitrarily small subintervals by rational interior points. The question in which intervals the roots are located can be answered as soon as we are in a position to decide how many roots lie between two given limits. By further subdividing the intervals containing roots we can approximate the real roots as close as we please.

A method to determine how many roots lie between two given limits, or how many roots there are altogether, is given by

Sturm's theorem. Starting from a given polynomial $X = f(x)$, *let the polynomials* $X_1, X_2, \ldots, X_r$ be determined as follows:

$$(1)\qquad \begin{array}{ll} X_1 = f'(x) & \text{(differentiation)},\\ \left.\begin{array}{l} X = Q_1 X_1 - X_2,\\ X_1 = Q_2 X_2 - X_3,\\ \cdots\cdots\cdots\cdots\\ X_{r-1} = Q_r X_r \end{array}\right\} & \begin{array}{l}\text{(Euclidean}\\ \text{algorithm)}\end{array} \end{array}$$

For every real number a *which is not a root of* $f(x)$ *let* $w(a)$ *be the number of variations in sign* [3] *in the number sequence*

$$X(a),\ X_1(a), \ldots,\ X_r(a)$$

in which all zeros are omitted. If b *and* c *are any numbers* ($b < c$) *for which* $f(x)$ *does not vanish, then the number of the various roots in the interval* $b \leqq x \leqq c$ (*multiple roots to be counted only once*) *is equal to*

$$w(b) - w(c).$$

The sequence of the polynomials $X, X_1, \ldots, X_r$ is called Sturm's *chain* [4] for $f(x)$. Thus, the theorem states that the number of zeros between b and c is given by the number of variations in sign in Sturm's chain which are lost in passing from b to c.

PROOF. Clearly, the last polynomial X_r of the chain is the g.c.d. of $X = f(x)$ and $X_1 = f'(x)$. If we divide all polynomials by X_r, we have removed the multiple linear factors from $f(x)$ without influencing the number of variations in sign at any point a which is not a root of X_r; for in the division all the signs of the terms of the chain have either remained unaltered, or all of them have been reversed. Thus, in the proof, we assume that the division has already been performed; then the last term of the chain is a constant distinct from zero. In general, the second term of the chain will no longer be the derivative of the first. In fact, if d, let us say, is a root of $f(x)$ of multiplicity l, we have

$$X = f(x) = (x-d)^l g(x), \quad g(d) \neq 0,$$
$$X_1 = f'(x) = l(x-d)^{l-1} g(x) + (x-d)^l g'(x).$$

Now the division by $(x-d)^{l-1}$ leads to two polynomials of the form

$$\overline{X} = (x-d)\, g(x),$$
$$\overline{X}_1 = l \cdot g(x) + (x-d)\, g'(x),$$

which may be divided by further factors for the other zeros $d', d'', \ldots$. We denote these modified polynomials of Sturm's chain again by $X = X_0, X_1, \ldots, X_r$.

---

[3] By the *sign* of a number c we mean the symbol $+$, $-$, or 0, according as c is positive, negative, or zero. If, in a succession of signs involving merely the symbols $+$ and $-$, a $+$ follows $-$, or vice versa, we speak of a *variation* in sign. If there are also zeros involved, they are omitted in counting the variations.

[4] Translator's note: Often called: Sturm's functions.

On this supposition, no two successive terms of the chain become zero at any point a. For if, let us say, $X_k(a)$ and $X_{k+1}(a)$ were both zero, we would infer from the equations (1) that $X_{k+2}(a), \ldots, X_r(a)$ are also zero, which is a contradiction, since $X_r = \text{constant}$ and $\neq 0$.

The roots of the polynomials of Sturm's chain divide the interval $b \leqq x \leqq c$ into subintervals. In such a subinterval neither X nor any X_k becomes zero, whence it follows by Weierstrass' Nullstellensatz that in the interior of such an interval all polynomials of Sturm's chain retain their signs so that the number $w(a)$ remains constant. It remains to be examined, how the number $w(a)$ changes at a point d where a polynomial of the chain vanishes.

Let d first be a root of $X_k (0 < k < r)$. According to the equation

$$X_{k-1} = Q_k X_k - X_{k+1}$$

the numbers $X_{k-1}(d)$ and $X_{k+1}(d)$ are necessarily of opposite sign. Thus, in the two adjacent subintervals X_{k-1} and X_{k+1} are of opposite sign. The sign of X_k ($+$, $-$, or zero) has no bearing on the number of changes of sign between X_{k-1} and X_{k+1}; there is always exactly one variation of sign. Therefore, the number $w(a)$ does not change at all at its passage through d.

Next, let d be a root of $f(x)$ so that, in accordance with the observation made at the outset, we have, say,

$$X = (x-d)\, g(x), \qquad g(d) \neq 0,$$
$$X_1 = l \cdot g(x) + (x-d)\, g'(x),$$

where l is an integer. The sign of X_1 at d and therefore in the two adjacent intervals is the same as that of $g(d)$, while that of X is equal to that of $(x-d)\, g(d)$ at every single point. Thus for $a < d$ we have a change of sign between $X(a)$ and $X_1(a)$, but for $a > d$ we have no change any more. Any other possible variations of sign in Sturm's chain are preserved at the passage through d, as has already been shown. Hence the number $w(a)$ decreases by 1 as a passes through d. This completes the proof of Sturm's Theorem.

If we wish to employ Sturm's Theorem for determining the total number of the various real roots of $f(x)$, the limits b and c must be, respectively, so small and so large that there are no more roots for either $x < b$ or for $x > c$. It suffices to take $b = -M$ and $c = M$. However, it is still more convenient to choose b and c so that all polynomials of Sturm's chain have no more zeros for $x < b$ or for $x > c$. Then their signs are determined by the signs of their leading coefficients: $a_0 x^m + a_1 x^{m-1} + \cdots$ has the sign of a_0 for very large x, and that of $(-1)^m a_0$ for very small (negative) x. In this method we may disregard the question as to how large b and c have to be: we merely compute the leading coefficients a_0 and degrees m of Sturm's polynomials.

EXERCISES. 1. Find the number of real roots of the polynomial

$$x^3 - 5x^2 + 8x - 8.$$

Between what successive integers do these roots lie?

2. If the last two polynomials X_{r-1}, X_r of Sturm's chain are of degree 1, 0, then the constant X_r (or its sign, which alone is of interest) can be found by substituting the root of X_{r-1} in $-X_{r-2}$.

3. If, in the computation of Sturm's chain, we encounter an X_k which changes its sign nowhere (e.g., a sum of squares), we may discontinue the chain with this X_k. Also, in every X_k we may always omit a factor which is positive everywhere, and continue the computation with the X_k thus modified.

4. The polynomial X_1 [a divisor of $f'(x)$] used in the proof of Sturm's Theorem surely changes its sign between two successive roots of $f(x)$. Give a proof, and derive from it that, between any two roots of $f(x)$, $f'(x)$ has at least one root (Rolle's Theorem).

5. Derive from Rolle's Theorem the *law of the mean in differential calculus* which states that, for $a < b$

$$\frac{f(b)-f(a)}{b-a} = f'(c)$$

for a suitable c so that $a < c < b$. [Take

$$f(x)-f(a)-\frac{f(b)-f(a)}{b-a}(x-a) = \varphi(x).]$$

6. If, in an interval $a \leqq x \leqq b$, where $f'(x) > 0$, $f(x)$ is an increasing function of x; if $f'(x) < 0$, $f(x)$ is a decreasing function.

7. A polynomial $f(x)$ has a maximum and a minimum value in every interval $a \leqq x \leqq b$, and the value of x for which the maximum is attained is either a root of $f'(x)$ or coincides with one of the endpoints a or b.

69. THE FIELD OF COMPLEX NUMBERS

If we adjoin to the field of real numbers P a root i of the polynomial x^2+1 which is irreducible in P, we obtain the *field of complex numbers* $\Omega = \mathsf{P}(i)$.

When speaking of "numbers," we mean only complex (and, in particular, real) numbers. *Algebraic numbers* are those numbers which are algebraic with respect to the rational number field Γ. It is now clear what is meant by algebraic number fields, real number fields, etc. By the theorems of Section 35, the algebraic numbers form a field A, which contains all algebraic number fields.

We now prove the following theorem:

In the field of complex numbers the equation $x^2 = a + bi$ $(a,b$ *real*$)$ *is always*

soluble, i.e., every number of the field has a "square root" in the field.

PROOF. A number $x=c+di$ (c, d real) has the demanded property if, and only if,

$$(c+di)^2=a+bi,$$

i.e., when the conditions

$$c^2-d^2=a, \qquad 2cd=b$$

are satisfied. From these equations follows: $(c^2+d^2)^2=a^2+b^2$; hence $c^2+d^2=\sqrt{a^2+b^2}$. From this and the first condition we find

$$c^2=\frac{a+\sqrt{a^2+b^2}}{2},$$

$$d^2=\frac{-a+\sqrt{a^2+b^2}}{2}.$$

The quantities on the right are actually $\geqq 0$. From them we can therefore determine c and d, except for the signs. Multiplying them, we get

$$4c^2d^2=-a^2+(a^2+b^2)=b^2;$$

Hence the signs of c and d can be determined so that the second condition

$$2cd=b$$

is satisfied.

It follows from what has been proved that, in the field of complex numbers, any quadratic equation

$$x^2+px+q=0$$

can be solved by writing it in the form

$$\left(x-\frac{p}{2}\right)^2=\frac{p^2}{4}-q.$$

The solution is

$$x=-\frac{p}{2}\pm w,$$

if w is any solution of the equation $w^2=\frac{p^2}{4}-q$.

The "*Fundamental Theorem of Algebra*," or more precisely, the fundamental theorem of the theory of complex numbers, states that not only every quadratic, but every non-constant polynomial $f(z)$ has a zero in the field Ω. The theory of complex functions furnishes the simplest proof of the Fundamental Theorem: Suppose the polynomial $f(z)$ has no complex zero; then

$$\frac{1}{f(z)}=\varphi(z)$$

would be a function regular in the entire z-plane which for $z\to\infty$ remains bounded (and even tends toward zero) and, therefore, would be a constant according to Liouville; but then $f(z)$ would also be a constant.

Assuming merely the basic fundamentals of the theory of functions, we can consider, instead of $\varphi(z)$, another rational function

$$\psi(z) = \frac{\varphi(z) - \varphi(0)}{z}$$

which, like $\varphi(z)$, is regular in the entire z-plane. The integral of this function over a circle K of radius R must therefore vanish:

$$\int_K \frac{\varphi(z) - \varphi(0)}{z}\, dz = \int_0^{2\pi} \{\varphi(Re^{i\vartheta}) - \varphi(0)\}\, i\, d\vartheta = 0.$$

Now, for sufficiently large R we have

$$|\varphi(Re^{i\vartheta})| < \varepsilon$$

$$\left|\int_0^{2\pi} \varphi(Re^{i\vartheta})\, i\, d\vartheta\right| < 2\pi\varepsilon,$$

furthermore,

$$\int_0^{2\pi} \varphi(0)\, i\, d\vartheta = 2\pi i\, \varphi(0)$$

so that

$$|2\pi i\, \varphi(0)| < 2\pi\varepsilon$$

$$\varphi(0) = 0$$

$$1 = f(0)\varphi(0) = 0,$$

giving a contradiction.

Gauss gave five proofs of the Fundamental Theorem. In Section 70 we shall become acquainted with Gauss' second proof, in which only the simplest properties of real and complex numbers are used; on the other hand, the algebraic devices used are quite intricate.[5]

By the *absolute value* $|\alpha|$ of the complex number $\alpha = a + bi$ we mean the real number

$$|\alpha| = \sqrt{a^2 + b^2} = \sqrt{\alpha\bar{\alpha}},$$

where $\bar{\alpha}$ is the conjugate complex number, i.e., the conjugate $a - bi$ with respect to the field of real numbers.

Obviously, $|\alpha| \geqq 0$ with $|\alpha| = 0$ only for $\alpha = 0$. Furthermore, we have $\sqrt{\alpha\beta\bar{\alpha}\bar{\beta}} = \sqrt{\alpha\bar{\alpha}} \cdot \sqrt{\beta\bar{\beta}}$, so that

$$|\alpha\beta| = |\alpha| \cdot |\beta|. \tag{1}$$

In order to prove the other relation

$$|\alpha + \beta| \leqq |\alpha| + |\beta| \tag{2}$$

we assume that, for the moment, the more special relation

$$|1 + \gamma| \leqq 1 + |\gamma| \tag{3}$$

is known. If $\alpha = 0$, (2) is trivial; but if $\alpha \neq 0$, we have

$$|\alpha + \beta| = |\alpha(1 + \alpha^{-1}\beta)| = |\alpha|\,|1 + \alpha^{-1}\beta|$$

$$\leqq |\alpha|(1 + |\alpha^{-1}\beta|) = |\alpha| + |\beta|.$$

[5] For another simple proof see: C. Jordan: *Cours d'Analyse I.*, 3me ed. p. 202. An intuitionistic proof is due to H. Weyl: *Math. Z. Vol 20 (1914)*, p. 142.

To prove (3) let $\gamma = a + bi$; then we have

$$|\gamma| = \sqrt{a^2 + b^2} \geqq \sqrt{a^2} = |a|,$$

$$|1+\gamma|^2 = (1+\gamma)(1+\bar{\gamma}) = 1 + \gamma + \bar{\gamma} + \gamma\bar{\gamma} =$$

$$= 1 + 2a + |\gamma|^2 \leqq 1 + 2|\gamma| + |\gamma|^2 = (1+|\gamma|)^2$$

so that

$$|1+\gamma| \leqq 1 + |\gamma|,$$

which proves (3).

70. ALGEBRAIC THEORY OF REAL FIELDS

One of the properties of ordered fields, especially of real number fields, is that a sum of squares vanishes in them only when the terms vanish individually, or, what is equivalent, that -1 is not expressible as a sum of squares.[6] In the field of complex numbers this is not true; for in it -1 is even a square. We shall see that this property is characteristic for the real algebraic number fields and their conjugate fields (in the field of all algebraic numbers), and it can be used for the algebraic construction of the field of real algebraic numbers including the conjugate fields. We lay down the following definition: [7]

A field will be called formally real if -1 *is not expressible in it as a sum of squares.*

A formally real field always has zero characteristic; for, in a field of characteristic p, -1 is always the sum of $p-1$ summands 1^2. Obviously, a subfield of a formally real field is formally real.

A field P *is called a real closed field if* P *is formally real but no proper algebraic extension of* P *is formally real.*

THEOREM 1. *Every real closed field can be ordered in one, and only one, way.*

Let P be a real closed field. We proceed to show the following properties:

If a is an element in P distinct from zero, then either a is itself a square, or $-a$ is a square, and these cases exclude one another. The sums of squares of elements in P are themselves squares.

From these properties Theorem 1 will follow at once. For by the stipulation $a > 0$, if a is a square and distinct from zero, we will obviously have defined an ordering of the field P, and this ordering is the only possible one, since, in any ordering, all squares must be $\geqq 0$.

---

[6] If, in any field, the element -1 is expressible as a sum Σa_ν^2, then $1^2 + \Sigma a_\nu^2 = 0$; thus, 0 is a sum of squares with bases not all vanishing. If, conversely, a relation $\Sigma b_\nu^2 = 0$ is given with at least one $b_\lambda \neq 0$, we can easily let this b_λ become 1 by dividing the sum by b_λ^2. Transposing the 1 to the other side, we obtain $-1 = \Sigma a_\nu^2$.

[7] Cf. E. Artin and O. Schreier: "Algebraische Konstruktion reeller Körper." *Abh. Math. Sem. Hamburg, Vol. 5 (1926), pp. 83-115.*

If γ is not the square of an element in P, then, if $\sqrt{\gamma}$ is a root of the polynomial $x^2-\gamma$, $\mathsf{P}(\sqrt{\gamma})$ is a proper algebraic extension of P, and is, therefore, not formally real. Therefore, an equation

$$-1=\sum_{\nu=1}^{n}(\alpha_\nu\sqrt{\gamma}+\beta_\nu)^2$$

or

$$-1=\gamma\sum_{\nu=1}^{n}\alpha_\nu^2+\sum_{\nu=1}^{n}\beta_\nu^2+2\sqrt{\gamma}\sum_{\nu=1}^{n}\alpha_\nu\beta_\nu$$

is valid, where the α_ν, β_ν belong to P. The last term must vanish; for, otherwise, $\sqrt{\gamma}$ would lie in P, contrary to the hypothesis. On the other hand, the first term on the right cannot vanish since, otherwise, P would not be formally real. From this we conclude first that γ is not expressible in P as a sum of squares; for, otherwise, -1 would be expressible as a sum of squares. This means: If γ is not a square, it cannot be the sum of squares. Or, turning this into a positive statement: Every sum of squares in P is a square in P.

We now obtain

$$-\gamma=\frac{1+\sum_{\nu=1}^{n}\beta_\nu^2}{\sum_{\nu=1}^{n}\alpha_\nu^2}.$$

Numerator and denominator of this expression are sums of squares and therefore themselves squares; hence $-\gamma=c^2$ with c lying in P. Consequently, at least one of the equations $\gamma=b^2$, $-\gamma=c^2$ is valid for every element γ in P; however, if $\gamma\neq 0$, both of them cannot hold since, otherwise, we would have $-1=\left(\frac{b}{c}\right)^2$, which is a contradiction.

On the basis of Theorem 1 we shall hereafter assume all real closed fields to be ordered.

THEOREM 2. *In a real closed field every polynomial of odd degree has at least one root.*

If the degree is 1, the theorem is trivial. We assume it to be true for all odd degrees $<n$; let $f(x)$ be a polynomial of odd degree n (>1). If $f(x)$ is reducible in the real closed field P, at least one irreducible factor is of odd degree $<n$, and, therefore, has a root in P. We proceed to show that the assumption that $f(x)$ is irreducible is absurd. Let, therefore, α be a symbolically adjoined root of $f(x)$. Then $\mathsf{P}(\alpha)$ would not be formally real; therefore we would have an equation

$$-1=\sum_{\nu=1}^{r}(\varphi_\nu(\alpha))^2, \tag{1}$$

where the $\varphi_\nu(x)$ are polynomials of at most degree $(n-1)$ with coefficient in P. From (1) we obtain an identity

$$-1 = \sum_{\nu=1}^{r} (\varphi_\nu(x))^2 + f(x)\, g(x). \tag{2}$$

The sum of the φ_ν^2 is of even degree, since the leading coefficients are squares and, therefore, cannot cancel out in the addition. Moreover, the degree is positive; for, otherwise, (1) would already contain a contradiction. Consequently $g(x)$ is of odd degree $\leqq n-2$; thus $g(x)$ definitely has one root a in P. However, substituting a in (2), we get

$$-1 = \sum_{\nu=1}^{r} (\varphi_\nu(a))^2,$$

which is a contradiction, since the $\varphi_\nu(a)$ lie in P.

THEOREM 3. *A real closed field is not algebraically closed. On the other hand, the field arising by the adjunction of i is algebraically closed.*[8]

The first part of the theorem is trivial; for the equation $x^2+1=0$ is insoluble in any formally real field.

The second part follows immediately from

THEOREM 3a. *If in an ordered field K every positive element possesses a square root and every polynomial of odd degree at least one root, then the field obtained by adjoining i is algebraically closed.*

First of all, we observe that every element has a square root in $\mathsf{K}(i)$, and that every quadratic equation is therefore soluble. The proof can be furnished by means of the same computation used in the field of complex numbers in Section 69.

In order to prove the algebraic closure of $\mathsf{K}(i)$ it suffices to show, by Section 62, that every polynomial $f(x)$ irreducible in K possesses a root in $\mathsf{K}(i)$. Let $f(x)$ be a polynomial of degree n without double roots, where $n=2^m q$, and q is odd. We employ the method of induction on m and assume that every polynomial without double roots and with coefficients in K, and whose degree is divisible by 2^{m-1}, but not by 2^m, possesses a root in $\mathsf{K}(i)$. By hypothesis, this is the case for $m=1$. Now, let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the roots of $f(x)$ in an extension of K. We choose c in K so that the values of the $\frac{n(n-1)}{2}$ expressions $\alpha_j\alpha_k + c(\alpha_j+\alpha_k)$ are all different[9] for $1 \leqq j < k \leqq n$. Since these expressions obviously satisfy an equation of degree $\frac{n(n-1)}{2}$ in K, at least one of them, say $\alpha_1\alpha_2 + c(\alpha_1+\alpha_2)$, lies in $\mathsf{K}(i)$ by hypothesis. But in consequence of the condition imposed on c we have (cf. Section 40)

$$\mathsf{K}(\alpha_1\alpha_2,\ \alpha_1+\alpha_2) = \mathsf{K}(\alpha_1\alpha_2 + c(\alpha_1+\alpha_2));$$

thus, we can find α_1 and α_2 by solving a quadratic equation in $\mathsf{K}(i)$.

[8] Here and in the following, i will always mean a root of x^2+1.

[9] This is possible because $f(x)$ was supposed to have no double roots.

At the same time it follows from Theorem 3a (in conjunction with Section 69) that the field of complex numbers is algebraically closed. This is the "Fundamental Theorem of Algebra" (cf. Section 69).

The converse of Theorem 3 is the following

THEOREM 4. *If a formally real field* K *can be closed algebraically by the adjunction of* i, K *is a real closed field.*

PROOF. There is no intermediate field between K and $\mathsf{K}(i)$, and so there is no algebraic extension of K, except K itself and $\mathsf{K}(i)$. $\mathsf{K}(i)$ is not formally real since -1 is a square in it. Hence K is a real closed field.

From Theorem 4 follows in particular that the field of real numbers is a real closed field.

The roots of an equation $f(x)=0$ with coefficients in a real closed field K lie in $\mathsf{K}(i)$ and, therefore, always occur in pairs of conjugate roots (with respect to K), insofar as they are not contained in K. If $a+bi$ is a root, $a-bi$ is its conjugate. By factoring $f(x)$ into linear factors and combining pairs of conjugate linear factors, we obtain a decomposition of $f(x)$ into linear and quadratic factors irreducible in K.

We are now in a position to extend "Weierstrass' Nullstellensatz" for polynomials (Section 68) to arbitrary real closed field.

THEOREM 5. *Let* $f(x)$ *be a polynomial with coefficients in a real closed field* P, *and let* a, b *be elements in* P, *for which* $f(a)<0$, $f(b)>0$. *Then there exists between* a *and* b *at least one element* c *in* P *such that* $f(c)=0$.

PROOF. As we have just seen, $f(x)$ resolves in P into linear and irreducible quadratic factors. An irreducible quadratic polynomial x^2+px+q is always positive in P; for it can be written in the form $\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)$; here the first term is always $\geqq 0$ and the second is positive because of the irreducibility assumed. Therefore, a change of sign of $f(x)$ can only be effected by a change of sign of a linear factor, i.e., of a root between a and b.

By virtue of this theorem, all deductions derived from Weierstrass' Nullstellensatz in Section 68, especially Sturm's Theorem on real roots, also hold for real closed fields.

We finally prove

THEOREM 6. *Let* K *be an ordered field and* $\overline{\mathsf{K}}$ *the field which arises from* K *through the adjunction of the square roots of all positive elements of* K. *Then* $\overline{\mathsf{K}}$ *is a formally real field.*

Obviously, it is sufficient to show that no equation of the form

$$-1=\sum_{\nu=1}^{n} c_\nu \xi_\nu^2 \tag{3}$$

exists, where the c_ν are positive elements in K, and the ξ_ν elements in $\overline{\mathsf{K}}$. Suppose that such an equation exists. In the ξ_ν we could actually have only a finite

number of square roots adjoined to K, such as $\sqrt{a_1}, \sqrt{a_2}, \ldots, \sqrt{a_r}$. From among all equations (3) let one be chosen for which r is as small as it can be. [Surely, $r \geqq 1$ since no equation of the form (3) exists in K.] ξ_ν can be represented in the form $\xi_\nu = \eta_\nu + \zeta_\nu \sqrt{a_r}$, where η_ν, ζ_ν lie in $\mathsf{K}(\sqrt{a_1}, \sqrt{a_2}, \ldots, \sqrt{a_{r-1}})$. Thus we would have

$$-1 = \sum_{\nu=1}^{n} c_\nu \eta_\nu^2 + \sum_{\nu=1}^{n} c_\nu a_r \zeta_\nu^2 + 2\sqrt{a_r} \sum_{\nu=1}^{n} c_\nu \eta_\nu \zeta_\nu . \tag{4}$$

If the last term in (4) vanishes, (4) is an equation of the same form as (3), but it contains less than r square roots. However, if it does not vanish, $\sqrt{a_r}$ would lie in $\mathsf{K}(\sqrt{a_1}, \ldots, \sqrt{a_{r-1}})$, and (3) would contain less than r square roots. Thus in either case our assumption leads to a contradiction.

EXERCISES. The field of algebraic numbers is algebraically closed, and the field of real algebraic numbers is a real closed field.

2. The algebraically closed algebraic extension field of the field Γ, which, by Section 62, is constructible by purely algebraic processes, is isomorphic with the field A of algebraic numbers.

3. Let P be a real number field, and Σ the field of real numbers algebraic with respect to P; then Σ is a real closed field.

4. If P is formally real, and t transcendental with respect to P, then $\mathsf{P}(t)$ is also a formally real field. [If $-1 = \Sigma \varphi_\nu(t)^2$, we substitute for t a suitable constant in P.]

71. EXISTENCE THEOREMS FOR FORMALLY REAL FIELDS

THEOREM 7. *Let* K *be a countable formally real field and* Ω *a countable algebraically closed field over* K; *then there exists (at least) one real closed field* P *between* K *and* Ω *such that* $\Omega = \mathsf{P}(i)$.

PROOF. Let the elements of Ω be $\omega_1, \omega_2, \ldots$. By the method of complete induction we now define a sequence of extension fields $\Sigma_1, \Sigma_2, \Sigma_3, \ldots$ of K as follows:

$$\Sigma_1 = \mathsf{K}$$

$$\Sigma_{n+1} = \Sigma_n(\omega_n) \quad \text{if } \Sigma_n(\omega_n) \text{ is a formally real field,}$$

otherwise $\Sigma_{n+1} = \Sigma_n$.

Finally, we define P as the union of all Σ_n.

By complete induction it follows at once that all fields Σ_n are formally real. But if all Σ_n are formally real, so is their union P; for if there exists a representation $-1 = \Sigma a_\nu^2$ in P, all the a_ν already belong to a single Σ_n.

Now let $\omega = \omega_n$ be an element of Ω which is not contained in P. Then ω_n is not contained in Σ_{n+1} either, so $\Sigma_n(\omega_n)$ is not formally real, nor is $\mathsf{P}(\omega)$. This is possible only when ω is algebraic over P; for a simple transcendental exten-

sion of a formally real field is itself formally real (Section 70, Ex. 4). Thus, every element of Ω is algebraic over P, i.e., Ω is algebraic over P. Furthermore, since we take for ω an arbitrary algebraic element of Ω not in P, no simple proper algebraic extension $\mathsf{P}(\omega)$ of P is formally real and therefore P is a real closed field. By Theorem 3 (Section 70), $\mathsf{P}(i)$ is algebraically closed and, therefore, identical with Ω. This completes the proof of the theorem.

NOTE. With slight modifications the above theorem and proof are also valid when Ω is not countable but well-ordered. Hence, on the basis of Zermelo's well-ordering theorem, we may state Theorem 7, even without assuming countability. For the most important applications, however, the countable case will suffice.

We proceed to state a few special cases and immediate consequences of Theorem 7.

THEOREM 7a. *To every countable formally real field* K *there exists (at least) one real closed algebraic extension.*

To prove this we merely choose the algebraically closed algebraic extension of K for Ω in Theorem 7.

THEOREM 7b. *Every countable formally real field can be ordered in (at least) one way.*

This follows immediately from Theorem 1 (Section 70) and Theorem 7a.

If, furthermore, Ω is any algebraically closed field of characteristic zero, and if we take the field of rational numbers for K in Theorem 7, we have

THEOREM 7c. *Every countable algebraically closed field* Ω *of zero characteristic contains (at least) one real closed subfield* P *such that* $\Omega = \mathsf{P}(i)$.

For ordered fields Theorem 7a can be restricted substantially:

THEOREM 8. *If* K *is a countable ordered field, then, except for equivalent extensions, there exists one, and only one, real closed algebraic extension* P *of* K, *whose ordering is a continuation of the ordering of* K. P *does not possess any automorphism, leaving the elements in* K *fixed, apart from the identical automorphism.*

PROOF. As in Theorem 6, we denote by $\bar{\mathsf{K}}$ the field which arises from K by the adjunction of the square roots of all positive elements of K. Let P be an algebraic real closed extension of $\bar{\mathsf{K}}$. By Theorem 7a, such an extension exists since $\bar{\mathsf{K}}$ is formally real. P is also algebraic with respect to K, and the ordering of P is a continuation of the ordering of K, since every positive element of K is a square in $\bar{\mathsf{K}}$ and, therefore, certainly in P. Thus, we have proved the existence of such a P.

For the uniqueness proof of P it is not necessary to assume the countability of K.

Let P^* be a second algebraic real closed extension of K, whose ordering continues that of K. Let $f(x)$ be a (not necessarily irreducible) polynomial with

coefficients in K. Sturm's Theorem allows us to determine already in K how many roots $f(x)$ has in P or in P*. We need merely investigate Sturm's chain for $f(x)=x^n+a_1x^{n-1}+\cdots+a_n$. Therefore, $f(x)$ has as many roots in P as it has in P*. In particular, every equation in K which has at least one root in P also has at least one root in P*, and vice versa. Let now $\alpha_1, \alpha_2, \ldots, \alpha_r$ be the roots of $f(x)$ in P, and $\beta_1^*, \beta_2^*, \ldots, \beta_r^*$ the roots of $f(x)$ in P*. Furthermore, let ξ be so chosen in P that $K(\xi)=K(\alpha_1, \ldots, \alpha_r)$ and that $F(x)=0$ is the irreducible equation for ξ in K. Thus $F(x)$ possesses the root ξ in P and, therefore, at least one root η^* in P*. $K(\xi)$ and $K(\eta^*)$ are equivalent extensions of K. Since $K(\xi)$ is generated by the r roots $\alpha_1, \ldots, \alpha_r$ of $f(x)$, $K(\eta^*)$ must be generated by the r roots of $f(x)$; now, $K(\eta^*)$ is a subfield of P*, so we have $K(\eta^*)=K(\beta_1^*, \ldots, \beta_r^*)$. Consequently $K(\alpha_1, \ldots, \alpha_r)$ and $K(\beta_1^*, \ldots, \beta_r^*)$ are equivalent extensions of K.

In order to show that P and P* are equivalent extensions of K we observe that an isomorphic mapping of P upon P* must necessarily preserve the ordering, since (by the proof of Theorem 1, Section 70) this ordering is determined by the property of any element of being or not being a square. We therefore define the following mapping σ of P upon P*. Let α be an element in P, $p(x)$ the irreducible polynomial in K having α as a root, and let $\alpha_1, \alpha_2 \ldots, \alpha_r$ be the roots of $p(x)$ in P, so numbered that $\alpha_1<\alpha_2<\cdots<\alpha_r$; in particular, let $\alpha=\alpha_k$. If $\alpha_1^*, \alpha_2^*, \ldots, \alpha_r^*$ are the roots of $p(x)$ in P*, and if $\alpha_1^*<\alpha_2^*\cdots<\alpha_r^*$, we put $\sigma(\alpha)=\alpha_k^*$. Obviously, σ is unique and leaves fixed the elements in K. It is to be proved that σ is an isomorphic mapping. For this purpose let $f(x)$ again be any polynomial in K, let $\gamma_1, \gamma_2, \ldots, \gamma_s$ be its roots in P, and $\gamma_1^*, \gamma_2^*, \ldots, \gamma_s^*$ those in P*. Furthermore, let $g(x)$ be the polynomial in K, whose roots are the square roots of the positive root differences of $f(x)$. Let $\delta_1, \delta_2, \ldots, \delta_t$ be the roots of $g(x)$ in P, and $\delta_1^*, \delta_2^*, \ldots, \delta_t^*$ those in P*. By the above proof, $\Lambda=K(\gamma_1, \ldots, \gamma_s, \delta_1, \ldots, \delta_t)$ and $\Lambda^*=K(\gamma_1^*, \ldots, \gamma_s^*, \delta_1^*, \ldots, \delta_t^*)$ are equivalent extensions of K. Thus, there exists an isomorphic mapping τ of Λ upon Λ^* which leaves each element of K fixed. τ associates a γ^* with every γ, and a δ^* with every δ. Let the notation be so that $\tau(\gamma_k^*)=\gamma_k$, $\tau(\delta^*)=\delta_k$. If $\gamma_k<\gamma_l$ (in P), we have $\gamma_l-\gamma_k=\delta_h^2$ for a certain index h, and so $\gamma_l^*-\gamma_k^*=\delta_h^{*2}$; hence $\gamma_k^*<\gamma_l^*$ (in P*). Thus, τ associates the roots of $f(x)$ in P with those in P* in increasing order. Since, in consequence, this is also true for the factors of $f(x)$ irreducible in K, we have $\tau(\gamma_k)=\sigma(\gamma_k)$ $(k=1, 2, \ldots, s)$. By taking care that two arbitrarily given elements α, β in P as well as $\alpha+\beta$ and $\alpha\cdot\beta$ occur among the roots of $f(x)$, we recognize that σ is an isomorphic mapping of P upon P*; it is the only one that leaves all the elements of K fixed. If we choose P*=P, the correctness of our assertion regarding the automorphisms of P becomes evident.

Since, by Section 66, the field of rational numbers Γ can be ordered in only one way, we infer from Theorem 8 at once:

THEOREM 8a. *Apart from isomorphic fields, there exists one, and only one, real closed algebraic field over* Γ.

For this field we may of course choose the field of real algebraic numbers in the ordinary sense (Section 67), which we obtain by singling out the algebraic numbers from among the real numbers. However, this is a transcendental detour, which can be avoided by the purely algebraic construction in Theorem 7 (in which we take $K=\Gamma$, and for Ω we take the algebraically closed algebraic extension field A over Γ). Thus, by a purely algebraic process, we construct the *field of real algebraic numbers*, which we denote by P. The field of all algebraic numbers is of the form $A=P(i)$.

As we shall see later, P is not the only real closed field in A but only one among an infinite number of equivalent ones.

THEOREM 9. *Every formally real countable algebraic extension field* K^* *of* Γ *is isomorphic with a subfield of* P, *and, therefore, with a real algebraic number field.*

PROOF. By Theorem 7a, we can always construct an algebraic real closed extension field P^* to K^*; by Theorem 8a, P^* necessarily turns out to be isomorphic with P. This proves the theorem.

A particular isomorphic mapping of K^* upon $K \subseteq P$ of course yields a particular ordering of K^* since all subfields K of P are ordered fields. Conversely, every ordering of K^* can be obtained in this manner, since the real closed extension field P^* constructed in the proof of Theorem 9 can, by Theorem 8, be so constructed that in its ordering that of K^* is preserved. Then under the isomorphism this ordering is carried into the (only possible) ordering of P.

If, in particular, we take for K^* a finite algebraic number field, which has only a finite number of isomorphisms in A, we obtain:

The number of isomorphisms which carry K^* *into a real algebraic number field is equal to the number of the various orderings which* K^* *is capable of* (*and, in particular, is equal to zero if* K^* *is not a formally real field*).

The fact that every formally real field in A can be extended to a real closed field $P^* \subset A$ makes us also recognize that there exists an infinite number of such fields P^* in A (though all these are isomorphic with each other, according to Theorem 8a); for all the fields $K_\zeta^* = \Gamma(\zeta\sqrt[n]{2})$, where n is an odd integer and ζ a n-th root of unity, are isomorphic with $\Gamma(\sqrt[n]{2})$, and so are formally real fields. Thus each of them leads to a real closed extension field P_ζ^*, and for a fixed n all these fields must be different, since an ordered field can only contain one n-th root of 2 (Section 66, Ex. 5). We can, however, choose the number r of these fields as large as we please.

EXERCISES. 1. Let ϑ be a root of the equation $x^4 - x - 1 = 0$ irreducible in Γ. In how many ways can the field $\Gamma(\vartheta)$ be ordered?

2. The field $\Gamma(t)$, where t is an indeterminate, can be ordered in an infinite number of ways, and the ordering can be Archimedean or non-Archimedean. t can be chosen infinitely large as well as infinitely small (cf. Section 66, Ex. 1).

3. How many roots does the polynomial $(z^2-t)^2-t^3$ possess in a real closed extension field of $\Gamma(t)$, if t is infinitely small? Where do these roots lie?

72. SUMS OF SQUARES

We now investigate the question as to which elements of a field K can be represented as sums of squares of elements in K.

For the present we may confine ourselves to formally real fields. For if K is not a formally real field, then -1 is a sum of squares, such as

$$-1=\sum_1^n \alpha_\nu^2.$$

If K has a characteristic other than 2, then, for an arbitrary element γ of K, the decomposition into $n+1$ squares follows:

$$\gamma=\left(\frac{1+\gamma}{2}\right)^2+(\sum \alpha_\nu^2)\left(\frac{1-\gamma}{2}\right)^2.$$

However, if K is of characteristic 2, the question is answered by the observation that every sum of squares is itself a square:

$$\sum \alpha_\nu^2=(\sum \alpha_\nu)^2.$$

That the sum and product of sums of squares are themselves sums of squares is readily seen; but even a quotient of sums of sqares is itself a sum of squares:

$$\frac{\alpha}{\beta}=\alpha\cdot\beta\cdot(\beta^{-1})^2.$$

We now prove the following theorem for formally real countable fields K:

If γ in K is not a sum of squares, there exists an ordering of K, in which γ turns out to be negative.

PROOF. Let γ not be a sum of squares. We first prove that $\mathsf{K}(\sqrt{-\gamma})$ is a formally real field. If $\sqrt{-\gamma}$ lies already in K, the proof is clear; if not, we conclude as follows: If we had

$$-1=\sum_1^n(\alpha_\nu\sqrt{-\gamma}+\beta_\nu)^2,$$

then, by the same reasoning as in the proof of Theorem 1 (Section 70), we would get

$$\gamma=\frac{1+\sum\beta_\nu^2}{\sum\alpha_\nu^2},$$

so γ would still be a sum of squares, which contradicts the hypothesis. Hence $\mathsf{K}(\sqrt{-\gamma})$ is a formally real field. Now, if $\mathsf{K}(\sqrt{-\gamma})$ is ordered according to

Theorem 7b (Section 71), the element, being a square, must turn out to be positive. This completes the proof of the theorem.

Applying the above to formally real algebraic number fields, we obtain the following theorem (noting that all possible orderings of such a field can be obtained through isomorphic mappings upon conjugate real number fields, according to Section 71):

An element γ of an algebraic number field K is the sum of squares if, and only if, the number γ is never carried into a negative number under the isomorphisms which carry K into its real conjugate fields.

If K is not a formally real field, this theorem is still valid, since in this case all numbers of K are sums of squares while there are no isomorphisms of the kind desired.

Such numbers of an algebraic number field K, which, in any isomorphic mapping of K upon a conjugate real number field, always go into positive numbers, are called *totally positive numbers in* K. If K does not possess any real conjugate fields, every number of K is to be called totally positive. The concept of total positiveness may be extended to any field K, those elements of K being totally positive which turn out to be positive in every possible ordering of K. If, in particular, no ordering of K exists, i.e., if K is not a formally real field, all numbers of K are totally positive. We may summarize the results of this section by stating that, *in a countable field of characteristics* $\neq 2$, *every totally positive element can be represented as a sum of squares.*

BIBLIOGRAPHY TO CHAPTER IX

Further theorems on the number of squares sufficient for the representation of totally positive numbers of a number field may be found in an article by E. Landau: "Über die Zerlegung total positiver Zahlen in Quadrate." *Göttinger Nachrichten. 1919,* p. 392. For the case of a field of rational functions see D. Hilbert: "Über die Darstellung definiter Formen als Summen von Formenquadraten." *Math. Ann. Vol. 32 (1888)*, pp. 342-350, and above all E. Artin: "Über die Zerlegung definiter Funktionen in Quadrate." *Abhandlungen aus dem Math. Seminar der Hamburgischen Universität, Vol. 5 (1926)*, pp. 100-115.

CHAPTER X

FIELDS WITH VALUATIONS

73. VALUATIONS

The construction of the field Ω_K (Section 67) to a given ordered field K does not make full use of the ordering of the field K, but only of the ordering of the absolute values $|a|$ of the field elements a. It is therefore reasonable to try to extend this construction to fields other than ordered fields for which a function $\varphi(a)$ with the properties of the absolute value exists.

A field is said to *have a valuation* if a function $\varphi(a)$ is defined for the elements a of K such that

1. $\varphi(a)$ is an element in an ordered field P.
2. $\varphi(a) > 0$ for $a \neq 0$; $\varphi(0) = 0$.
3. $\varphi(ab) = \varphi(a)\varphi(b)$.
4. $\varphi(a+b) \leqq \varphi(a) + \varphi(b)$.

From 2. and 3. follows at once

$$\varphi(1) = 1, \quad \varphi(-1) = 1, \quad \varphi(a) = \varphi(-a).$$

From 4. follows

$$\varphi(c) - \varphi(a) \leqq \varphi(c-a),$$

also

$$\varphi(a) - \varphi(c) \leqq \varphi(c-a),$$

and so finally

$$|\varphi(c) - \varphi(a)| \leqq \varphi(c-a).$$

The properties 1. to 4. are fulfilled if K is ordered and if we put $\varphi(a) = |a|$. Furthermore, every field has the "trivial" valuation $\varphi(a) = 1$ for $a \neq 0$, $\varphi(0) = 0$. But there also exist entirely different types of valuations. Let Γ be the field of rationals. If p is a fixed prime, and if we write every rational number $a \neq 0$ in the form

$$a = \frac{s}{t} p^n,$$

where s and t are integers not divisible by p, then

$$\varphi_p(a) = p^{-n}, \quad \varphi_p(0) = 0$$

defines a valuation of Γ. 1. to 3. can be verified easily. Instead of 4. the stronger inequality

(1) $$\varphi_p(a+b) \leqq \max(\varphi_p(a), \varphi_p(b))$$

is valid. For if

$$a = \frac{s}{t} p^n, \quad b = \frac{u}{v} p^m, \text{ where } s, t, u, v \text{ are prime to } p,$$

and, let us say, $\varphi_p(b) \geqq \varphi_p(a)$, i.e., $n \geqq m$, then

$$a+b = \frac{s v p^{n-m} + t u}{t v} p^m,$$

and thus we get

$$\varphi_p(a+b) = \rho^{-m'} \text{ where } m' \geqq m$$

so that

$$\varphi_p(a+b) \leqq \varphi_p(b).$$

This is the *p-adic valuation of* Γ.

The p-adic valuation can be readily generalized. Let $\mathfrak{o}$ be an integral domain, K its quotient field, and $\mathfrak{p}$ a prime ideal of $\mathfrak{o}$ with the following properties: *A. All powers* $\mathfrak{p}, \mathfrak{p}^2, \ldots$ *are distinct, and their intersection is empty. B. If an element* a *is divisible exactly by* $\mathfrak{p}^\alpha$, *i.e. by* $\mathfrak{p}^\alpha$ *but not by* $\mathfrak{p}^{\alpha+1}$, *and, similarly, if* b *is divisible exactly by* $\mathfrak{p}^\beta$, *then* ab *is divisible exactly by* $\mathfrak{p}^{\alpha+\beta}$. Here $\mathfrak{p}^\alpha$ denotes the totality of all sums $\sum_\nu p_{\nu 1} p_{\nu 2} \cdots p_{\nu\alpha}$, where all $p_{\nu\varkappa}$ are elements of $\mathfrak{p}$. In particular, we have $\mathfrak{p}^1 = \mathfrak{p}$, $\mathfrak{p}^0 = \mathfrak{o}$. Now, if an element a is divisible exactly by $\mathfrak{p}^\alpha$, we define

$$\varphi(a) = e^{-\alpha} \quad \text{and} \quad \varphi(0) = 0,$$

where e is any real number > 1. Then the valuation $\varphi(a)$ is defined for the elements of $\mathfrak{o}$ and has the properties 1. to 4.

If a valuation for the elements of an integral domain is defined, then, by defining

$$\varphi\left(\frac{a}{b}\right) = \frac{\varphi(a)}{\varphi(b)},$$

it can immediately be extended to the elements of the quotient field. The definition is unique; for

$$\frac{a}{b} = \frac{c}{d} \quad \text{or} \quad ad = bc$$

implies

$$\varphi(a)\varphi(d) = \varphi(b)\varphi(c) \quad \text{or} \quad \frac{\varphi(a)}{\varphi(b)} = \frac{\varphi(c)}{\varphi(d)}.$$

Again, the valuation $\varphi\left(\frac{a}{b}\right)$ has the properties 1. to 4. The first three are evident. Property 4. is obtained thus:

$$\varphi\left(\frac{a}{b} + \frac{c}{d}\right) = \frac{\varphi(ad+bc)}{\varphi(bd)} \leqq \frac{\varphi(ad)+\varphi(bc)}{\varphi(bd)} = \varphi\left(\frac{a}{b}\right) + \varphi\left(\frac{c}{d}\right).$$

In this way we immediately obtain a valuation of the quotient field K from the valuation of the integral domain $\mathfrak{o}$, defined by the prime ideal $\mathfrak{p}$. The valuation of the quotient field K is called the $\mathfrak{p}$-adic valuation of K.

The properties *A.*, *B.* are fulfilled for many prime ideals. For example, in rings with a unique factorization all prime principal ideals possess these properties. In the polynomial ring $\Delta[x_1, \ldots, x_n]$ the ideal

$$\mathfrak{p} = (x_1, \ldots, x_n)$$

also has the properties *A.* and *B.* The corresponding valuation $\varphi(f)$ is $e^{-\alpha}$, where α is the degree of the terms of lowest degree which occur in the polynomial f.

EXERCISES. 1. Let the requirement in the valuation definition that $\varphi(a)$ shall not be negative be dropped. To prove: If there exists an element c in K such that $\varphi(c) < 0$, then $a \to \varphi(a)$ is an isomorphic mapping of K upon a subfield of the valued field P. [Prove that the equality sign is valid in 4., by taking into account the inequality for $\varphi(ac + bc)$ corresponding to 4.]

2. In all $\mathfrak{p}$-adic valuations (1) holds.

The most important investigations of fields with a valuation concern the case, in which the ordering of the field P is Archimedean. In this case, P may, by Section 67, Ex. 3, be imbedded in the field of real numbers. Therefore, we now suppose that the values $\varphi(a)$ are real numbers. We assume that the reader is familiar with the (natural) logarithms of the real numbers and their most elementary properties, as well as with the powers α^β of a positive number α with arbitrary real exponents.

Moreover, we make use of the following lemma on real numbers:

If α, β, γ are positive real numbers, and if

$$\gamma^\nu \leqq \alpha\nu + \beta$$

holds for every natural number ν, then $\gamma \leqq 1$.

PROOF. Suppose $\gamma = 1 + \delta$, $\delta > 0$. Then, for $\nu \geqq 2$, we would have

$$\gamma^\nu = (1+\delta)^\nu = 1 + \nu\delta + \tfrac{1}{2}\nu(\nu-1)\delta^2 + \cdots > \nu\delta + \tfrac{1}{2}\nu(\nu-1)\delta^2;$$

yet for sufficiently large ν we have in any event

$$\nu\delta > \beta \quad \text{and} \quad \tfrac{1}{2}(\nu-1)\delta^2 > \alpha,$$

so that we would have

$$\gamma^\nu > \beta + \alpha\nu,$$

contrary to the hypothesis.

A real valuation $\varphi(a)$ of a field K is called *non-Archimedean* if, for all natural multiples $n = 1 + 1 + 1 \cdots + 1$ of one, the condition

$$\varphi(n) \leqq 1$$

is valid. The *p*-adic valuation of the field of rationals is non-Archimedean. The field of values P is Archimedean, but that is an entirely different question.

The valuation φ *of* K *is non-Archimedean if, and only if, instead of* 4, *the stronger inequality*

4'. $$\varphi(a+b) \leqq \max(\varphi(a), \varphi(b))$$

is valid.

PROOF. 1. If 4'. is valid, we have

$$\varphi(n) \leqq \max(\cdots, \varphi(1), \cdots) = 1$$

for $n = 1+1+\cdots+1$.

2. If φ is non-Archimedean, then, for $\nu = 1, 2, 3, \ldots$, we have

$$(\varphi(a+b))^\nu = \varphi((a+b)^\nu) = \varphi(a^\nu + \binom{\nu}{1} a^{\nu-1}b + \cdots + b^\nu)$$
$$\leqq \varphi(a)^\nu + \varphi(a)^{\nu-1}\varphi(b) + \cdots + \varphi(b)^\nu \leqq (\nu+1)M^\nu,$$

$M = \max(\varphi(a), \varphi(b))$. By the lemma it follows that

$$\frac{\varphi(a+b)}{M} \leqq 1, \text{ or } \varphi(a+b) \leqq M,$$

i.e., 4'. is valid.

In the sequel we shall regard the inequality 4'. as a characteristic of a non-Archimedean valuation even when the field of values does not consist of real numbers. According to Krull,[1] an arbitrary ordered Abelian group can be taken as the domain of values, since the postulates 1., 2., 3., 4'. do not involve any addition of the values but only their multiplication and comparison.

It is often useful (and customary in the literature) to introduce another notation for non-Archimedean valuations. Instead of the real value $\varphi(a)$ one considers the *exponent* $w(a) = -\log \varphi(a)$. The defining properties of the valuation in terms of exponents are:

1. For every $a \neq 0$, $w(a)$ is a real number.
2. $w(0)$ is the symbol ∞.
3. $w(ab) = w(a) + w(b)$.
4. $w(a+b) \geqq \min(w(a), w(b))$.

In this case we speak of an *exponential valuation.* The introduction of exponents is made possible by the fact that, in 4'., the addition of the values $\varphi(a)$ need not be performed. The formation of logarithms inverts the ordering and transmutes the multiplication into addition.

EXAMPLE: Let the elements of the field K be unique analytic (meromorphic) functions defined in a region in the z-plane, or more generally, on a Riemann surface. We choose a certain point P on the Riemann surface and adopt the following definition: The value $w(a)$ of a function a shall be equal to α if the function has a zero of order α at P; it shall be equal to zero if the function has a finite value at P, distinct from zero, and the value shall be equal to $-\alpha$ if the function has a pole of order α at P. Then the properties 1. to 4. hold. Thus, to every point P belongs a valuation of the field K. By means of this example we can

[1] Krull, W: "Allgemeine Bewertungstheorie." *J. reine angew. Math. Volume 167 (1932)* pp. 160-196.

guess the significance of the valuation theory for the theory of algebraic functions of a complex variable. This example also explains why the valuations of a field K are occasionally called "places" (*Stellen*).

There are two types of exponential valuations: *Discrete* valuations, which are characterized by the fact that there exists a smallest positive value $w(a)$ such that all values $w(a)$ involved are multiples of it (cf. the above example), and *non-discrete* valuations, in which the $w(a)$ involved become arbitrarily small and tend towards zero. Since the multiples of a value $w(a)$ are themselves values $nw(a) = w(a^n)$, the values $w(a)$ in the non-discrete case are everywhere dense in the set of real numbers.

The p-adic valuation of the rational number field is discrete, and so are all p-adic valuations.

In a field K with an exponential valuation the elements a for which $w(a) \geqq 0$ form a ring $\mathfrak{J}$. For $w(a) \geqq 0$ and $w(b) \geqq 0$ imply $w(a \pm b) \geqq \min(w(a), w(b)) \geqq 0$ and $w(ab) = w(a) + w(b) \geqq 0$. The totality $\mathfrak{p}$ of all the elements a of K with $w(a) > 0$ is a prime ideal of $\mathfrak{J}$; for, first, from $w(a) > 0$, $w(b) > 0$ follows $w(a \pm b) \geqq \min(w(a), w(b)) > 0$; hence $\mathfrak{p}$ is a module. Secondly, from $a \in \mathfrak{p}$, i.e., $w(a) > 0$ and $w(c) \geqq 0$ follows $w(ca) = w(c) + w(a) > 0$; hence $\mathfrak{p}$ is an ideal. Thirdly, from $ab \equiv 0 \pmod{\mathfrak{p}}$, i.e., from $w(ab) = w(a) + w(b) > 0$ follows that at least one of the two numbers $w(a)$ and $w(b)$ is positive, so that at least one of the two elements a and b is divisible by $\mathfrak{p}$; hence $\mathfrak{p}$ is prime.

$\mathfrak{J}$ is called the *valuation ring* to the valuation w. The elements in $\mathfrak{J}$ are called *integral* (with respect to the valuation). An element a is called *divisible* by b (with respect to the valuation w) if $\frac{a}{b}$ is integral, or if $w(a) \geqq w(b)$.

The elements a for which $w(a) = 0$ are the units of the ring $\mathfrak{J}$. Since all elements of $\mathfrak{J}$ not belonging to $\mathfrak{p}$ are units of $\mathfrak{J}$, $\mathfrak{p}$ is a maximal ideal of $\mathfrak{J}$. Hence the residue class ring $\mathfrak{J}/\mathfrak{p}$ is a field, called the *residue class field* of the valuation. If the field K is of characteristic p, then, obviously, the residue class field is also of characteristic p. However, if K is of characteristic zero, the residue class field may be either of characteristic zero (case of equal characteristic), or it may have a prime number characteristic (case of different characteristics). The p-adic valuations are typical examples for the case of different characteristics. An example for the case of equal characteristic is given by the field of rational functions of one variable if the exponential value $w(a)$ of a rational function a is equated to the degree of the denominator minus the degree of the numerator. The p-adic valuations (see above) defined by ideals of the polynomial ring $K[x_1, \ldots, x_n]$ constitute a case of equal characteristic as well.

For further information regarding these conceptions, including a complete classification of all valuations, the reader is referred to the papers by H. Hasse, F. K. Schmidt, O. Teichmüller, and A. Witt.[2]

[2] Witt, E.: *J. reine u. angewandte Math. Vol. 176 (1936)* pp. 126-140, and the bibliography quoted there.

EXERCISES. 3. Show that in $\mathfrak{J}$ every ideal is either the set of all a for which $w(a) > \delta$, or the set of all a for which $w(a) \geqq \delta$, where δ is a non-negative real number. In a discrete valuation we can confine ourselves to the case $\geqq$ and take for δ a number which actually occurs in the set of values. In a non-discrete valuation δ is uniquely determined by the ideal.

4. In a discrete valuation all ideals of $\mathfrak{J}$ are powers of $\mathfrak{p}$, whereas in a non-discrete valuation all powers of $\mathfrak{p}$ are equal to $\mathfrak{p}$.

74. COMPLETE FIELD EXTENSIONS

An extension field Ω_{K} with a valuation, for which *Cauchy's convergence theorem* holds, can be constructed to every field K with a valuation, according to exactly the same method as that in Section 67. For this purpose it is necessary to construct beforehand, by the construction of Section 67, to the ordered field P the ordered extension field Ω, which will serve as the field of values for Ω_{K}. We now define fundamental sequences $\{a_\nu\}$ in K by the property

$$\varphi(a_p - a_q) < \varepsilon \quad \text{for} \quad p > n(\varepsilon), \quad q > n(\varepsilon),$$

where ε is an arbitrary positive quantity in P. The residue class field Ω_{K} is obtained from the ring of fundamental sequences just as in Section 67; all proofs are valid literally. The only difference is that Ω_{K}, like K, is not ordered but is merely a field with a valuation. The valuation of Ω_{K} is defined thus: If α is defined by the fundamental sequence $\{a_\nu\}$, then, by the already proven inequality

$$|\varphi(a_\nu) - \varphi(a_\mu)| \leqq \varphi(a_\nu - a_\mu),$$

the values $\varphi(a_\nu)$ in P also form a fundamental sequence which thus has a limit ω in $\Omega_{\mathfrak{p}}$. Then we define

$$\varphi(\alpha) = \omega.$$

All fundamental sequences with the same limit α define the same value $\varphi(\alpha)$, which satisfies the requirements 1.–4. *Ω_{K} is called the complete extension of K relative to the valuation φ.*

For the exponential valuations, Cauchy's convergence criterion assumes the following simple form:

If the field K is complete relative to the exponential valuation w, then a sequence $\{a_\nu\}$ is convergent as soon as

$$\lim_{\nu \to \infty} w(a_{\nu+1} - a_\nu) = \infty;$$

for in our case Cauchy's general convergence criterion runs as follows: $\{a_\nu\}$ is convergent if $w(a_{\nu+k} - a_\nu) > \mathrm{M}$ for all k, as soon as $\nu > n(\mathrm{M})$. However, since $w(a_{\nu+k} - a_\nu) \geqq \min(w(a_{\nu+k} - a_{\nu+k-1}), \ldots, w(a_{\nu+1} - a_\nu))$, the condition

$$\lim_{\nu \to \infty} w(a_{\nu+1} - a_\nu) = \infty$$

is sufficient for convergence.

An alternative form of this criterion is the following: *For an infinite series* $a_1+a_2+a_3+\cdots$ *to be convergent, it is necessary and sufficient that*

$$\lim_{\nu\to\infty} w(a_\nu)=\infty .$$

In the following we confine ourselves to the case in which all values are real numbers. Then the field P is a real number field, and Ω is the field of real numbers; we may assume $\mathsf{P}=\Omega$ at the outset.

If we start with the valuation of the field Γ of rational numbers by ordinary absolute values $\varphi(a)=|a|$, we naturally obtain the field of real numbers as the complete extension. However, if we start with the p-adic valuation of Γ, the complete extension we obtain will be the *field* Ω_p *of Hensel's p-adic numbers.*

The elements of the field Ω_p, i.e., the p-adic numbers, can be represented in a more convenient way than by arbitrary fundamental sequences. In order to show this, let us consider for $\lambda=0, 1, 2, 3, \ldots$ the module $\mathfrak{M}_\lambda$, consisting of those rational numbers whose numerators are divisble by p^λ, and whose denominators are not divisible by p so that $\varphi(a)\leqq p^{-\lambda}$. Two rational numbers will be called congruent (mod p^λ) if their difference belongs to $\mathfrak{M}_\lambda$. Now, if $\{r_\mu\}$ is a p-adic fundamental sequence of rational numbers, then, for every λ, starting from a certain $n=n(\lambda)$, we have

$$\varphi(r_\mu-r_\nu)\leqq p^{-\lambda} \quad \text{for} \quad \mu>n(\lambda),\ \nu>n(\lambda),$$

i.e.

$$r_\mu\equiv r_\nu \pmod{p^\lambda}.$$

Thus all numbers r_μ, where $\mu>n(\lambda)$, belong to a single residue class $\mathfrak{R}_\lambda$ modulo $\mathfrak{M}_\lambda$. Therefore, the fundamental sequence $\{r_\mu\}$ defines a sequence of residue classes

$$\mathfrak{R}_0\supset\mathfrak{R}_1\supset\mathfrak{R}_2\supset\mathfrak{R}_3\supset\mathfrak{R}_4\supset\ldots,$$

which are nested in the manner indicated. Conversely, any sequence $\{r_1, r_2, \ldots\}$ which defines a sequence $\{\mathfrak{R}_\lambda\}$ of nested residue classes $\mathfrak{R}_\lambda$ modulo $\mathfrak{M}$ in the indicated manner, so that

$$r_\mu \quad \text{in } \mathfrak{R}_\lambda \text{ for all } \mu>n(\lambda)$$

is always a fundamental sequence.

If, in particular, $\{r_\mu\}$ is a null sequence, $\mathfrak{R}_\lambda=\mathfrak{M}_\lambda$ becomes the null residue class. If we add two fundamental sequences $\{r_\mu\}+\{s_\mu\}=\{r_\mu+s_\mu\}$, we have to add the respective residue class sequences: $\{\mathfrak{R}_\lambda+\mathfrak{S}_\lambda\}$. If, in particular, we add a null sequence to a fundamental sequence, the respective residue class sequence remains unchanged. If, conversely, two sequences $\{r_\mu\}$ and $\{s_\mu\}$ belong to the same residue class sequence $\{\mathfrak{R}_\lambda\}$, their difference is a null sequence. Therefore, *a residue class sequence* $\{\mathfrak{R}_\lambda\}$ *of the above kind corresponds biuniquely to every p-adic number* $\alpha=\lim r_\nu$.

This representation of the p-adic numbers by residue class sequences is the convenient representation alluded to. In order to return to a (particular) funda-

mental sequence from a residue class representation of a p-adic number α, we need only choose an r'_λ from every residue class $\mathfrak{R}_\lambda$; then $\alpha = \lim r'_\lambda$. We may also represent α as an infinite sum by taking

$$r'_1 = s_0, \quad r'_{\lambda+1} - r'_\lambda = s_\lambda p^\lambda;$$

then we get

$$r_{\lambda+1} = s_0 + s_1 p + s_2 p^2 + \cdots + s_\lambda p^\lambda,$$

and so

$$\alpha = \lim_{\lambda \to \infty} \sum_{\nu=0}^{\lambda} s_\nu p^\nu = \sum_{\nu=0}^{\infty} s_\nu p^\nu. \tag{1}$$

Here $s_1, s_2 \ldots$ are rational numbers with denominators not divisible by p.

A p-adic limit of ordinary integers is called a *p-adic integer*. For the residue classes $\mathfrak{R}_0, \mathfrak{R}_1, \ldots$ this means that in each of them an integer must occur. For the null residue class $\mathfrak{R}_0$, this implies that $\mathfrak{R}_0$ is the totality of rational numbers with denominators not divisible by p. This condition is sufficient for $\mathfrak{R}_0$ to be a p-adic integer: If $\mathfrak{R}_0$ is the null residue class modulo $\mathfrak{M}_0$, all residue classes $\mathfrak{R}_1, \mathfrak{R}_2, \ldots$ contain integers; for $\mathfrak{R}_\lambda$ is contained in $\mathfrak{R}_0$ and, therefore, consists only of numbers r/s with $s \not\equiv 0 \pmod{p}$. Solving the congruence

$$sx \equiv r \pmod{p^\lambda},$$

we get

$$x - \frac{r}{s} = \frac{sx - r}{s} \equiv 0 \pmod{\mathfrak{M}_\lambda};$$

hence the number x belongs to the residue class $\mathfrak{R}_\lambda$.

Consequently, if α is a p-adic integer, r'_λ and all s_ν can be chosen as ordinary integers in the series representation (1). Thus, (1) is a power series in p with integral coefficients. Any such power series converges in the sense of the p-adic valuation and represents a p-adic integer.

Every p-adic number α with a residue class representation $\{\mathfrak{R}_0, \mathfrak{R}_1, \ldots\}$ can be transmuted into a p-adic integer by multiplying it by a power of p. For if r'_0 is an element of the residue class $\mathfrak{R}_0$, then, if we multiply r'_0 by a power p^m of p, the denominator of $p^m r'_0$ will no longer contain a factor p so that r'_0 belongs to the null residue class modulo $\mathfrak{M}_0$. If we now expand the p-adic integer $p^m \alpha$ in a power series (1) with integral $s_0, s_1, \ldots$, we obtain for α a representation with a finite number of negative exponents:

$$\alpha = a_{-m} p^{-m} + a_{-m+1} p^{-m+1} + \cdots + a_0 + a_1 p + a_2 p^2 + \cdots. \tag{2}$$

The representation (1) of the p-adic integer α can be normed by always choosing for r'_λ the smallest non-negative integer in the residue class $\mathfrak{R}_\lambda$. Then all numbers s_ν satisfy the condition $0 \leqq s_\nu < p$. Now, if we again pass from (1) to (2), then we obtain *for every p-adic number a uniquely determined expansion* (2) *with* $0 \leqq a_\nu < p$.

From the representation (2) of the p-adic numbers follows at once that the valuation ring of the p-adic valuation in Ω_p consists exactly of all p-adic integers so

that the term "integral," introduced in Section 73, is in conformity with the notion of "p-adic integers."

From the $\mathfrak{p}$-adic valuation of a field K defined, according to Section 73, by a prime ideal $\mathfrak{p}$ of an integral domain $\mathfrak{o}$ we also obtain a perfect $\mathfrak{p}$-*adic field* $\Omega_{\mathfrak{p}}$, the generalization of Hensel's p-adic field. For example, if $\mathfrak{p}$ is the ideal $(x-c)$ in the polynomial domain $\Delta[x]$, $\Omega_{\mathfrak{p}}$ becomes the ring of all power series

$$\alpha = a_{-m}(x-c)^{-m} + \cdots + a_0 + a_1(x-c) + a_2(x-c)^2 + \cdots \tag{3}$$

with constant a_ν in Δ (proof as above). The power series *always* converges in the sense of the $\mathfrak{p}$-adic valuation, no matter what coefficients a_ν are chosen. The expressions (3) are called *formal power series in* $(x-c)$.

EXERCISES. 1. Express -1 and $\frac{1}{2}$ as 3-adic normed power series.

2. An equation $f(\xi)=0$, where f is an integral polynomial, is soluble in the field Ω_p if, and only if, the congruence

$$f(\xi) \equiv 0 \pmod{p^n}$$

has a rational solution ξ for every natural number n.

3. Are the equations

$$x^2 = -1, \quad x^2 = 3, \quad x^2 = 7$$

soluble in the field Ω_3?

4. Let $\mathfrak{p}$ be the ideal $(x_1, x_2, \ldots, x_n)$ in the polynomial domain

$$\Delta[x_1, x_2, \ldots, x_n].$$

Show that the congruence

$$fg \equiv 1 \pmod{\mathfrak{p}}$$

has a solution for a given $f \not\equiv 0 \pmod{\mathfrak{p}}$. Show, by means of this proof, that the $\mathfrak{p}$-adic field $\Omega_{\mathfrak{p}}$ consists of all quotients of formal power series

$$h_0 + h_1 + h_2 + \cdots$$

in which h_k is a homogeneous form of degree k in $x_1, \ldots, x_n$.

It is possible that two different valuations φ and ψ of a field K lead to the same complete extension field Ω. Obviously, this is the case if, and only if, every sequence $\{a_\nu\}$ of K, which is a null sequence for φ is also a null sequence for ψ, and vice versa. In this case, i.e., when $\lim_{\nu\to\infty} \varphi(a_\nu) = 0$ and $\lim_{\nu\to\infty} \psi(a_\nu) = 0$ mean the same, the two valuations φ and ψ will be called *equivalent.*

We can form an infinite number of equivalent valuations to the valuation $\varphi(a) = |a|$ of the field of complex numbers by ordinary absolute values if we take $\varphi(a) = |a|^\varrho$, where ϱ is a fixed positive real number which is not greater than 1. Conditions 1. to 3. are satisfied in a trivial fashion. 4. follows from $|a+b| \leqq |a| + |b|$ by means of the inequality $\varepsilon^\varrho + \delta^\varrho \leqq (\varepsilon+\delta)^\varrho$, which is valid for any two real numbers $\varepsilon \geqq 0$, $\delta \geqq 0$ and $0 < \varrho \leqq 1$.[3]

[3] Cf., e.g., Hardy-Littlewood-Polya: *Inequalities.* Cambridge 1934. Chapter II.

Every valuation $\psi(a)=\varphi_p(a)^\sigma$, where σ is any fixed positive number, is equivalent to the p-adic valuation $\varphi_p(a)$ of the field of rational numbers.

If φ and ψ are two equivalent valuations of a field K, *then ψ is a power of φ, i.e., there exists a fixed positive number ε such that $\psi(a)=\varphi(a)^\varepsilon$ for all a in* K.

PROOF. If $\varphi(a)<\varphi(b)$, we also have $\psi(a)<\psi(b)$, and vice versa; for $\varphi(a)<\varphi(b)$ implies $\varphi(a/b)<1$; so for $n\to\infty$, $(a/b)^n$ converges to zero in the sense of the valuation φ. Therefore, $(a/b)^n$ converges to zero for ψ as well. But this means that $\psi(a/b)<1$ or $\psi(a)<\psi(b)$. Let p now be any fixed element of K for which $\varphi(p)>1$. Then we also have $\psi(p)>1$. Let a be an arbitrary element of K, and let $\varphi(a)=\varphi(p)^\delta$, $\psi(a)=\psi(p)^{\delta'}$. We wish to show that $\delta=\delta'$. Let n and m be positive integers such that $n/m\leqq\delta$. Then we have

$$\varphi(p)^{n/m}\leqq\varphi(p)^\delta=\varphi(a),\quad\text{hence}\quad\varphi(p^n)\leqq\varphi(a^m).$$

This implies

$$\psi(p^n)\leqq\psi(a^m),\quad\psi(p)^{n/m}\leqq\psi(a)=\psi(p)^{\delta'},\quad n/m\leqq\delta'.$$

Since the least upper bound of all fractions $n/m\leqq\delta$ is exactly δ, it follows that $\delta\leqq\delta'$, and similarly $\delta'\leqq\delta$ so that $\delta=\delta'$. Now, since $\varepsilon=\dfrac{\log\psi(p)}{\log\varphi(p)}$ is a fixed positive number independent of a, and $\delta=\delta'$, we have

$$\log\psi(a)=\delta'\log\psi(p)=\delta\log\psi(p)=\delta\varepsilon\log\varphi(p)=\varepsilon\log\varphi(a)$$

for all a; hence

$$\psi(a)=\varphi(a)^\varepsilon.$$

If K is a field with the valuation φ, and K′ a field isomorphic with K and with the valuation ψ, then an isomorphism between K and K′ is called *topological* if it always maps a φ-null sequence of K upon a ψ-null sequence of K′, and vice versa. In this case the fields K and K′ are called *topologically isomorphic*. In a topological isomorphism convergent sequences and fundamental sequences correspond. From this follows readily:

Topologically isomorphic fields K *and* K′ *have topologically isomorphic perfect extensions Ω_K and $\Omega_{K'}$.*

EXERCISE 5. Show that no two of the valuations of the field of rationals known to us, namely the valuation according to absolute value and the p-adic valuations, are equivalent.

75. VALUATIONS OF THE FIELD OF RATIONAL NUMBERS. ARCHIMEDEAN VALUATIONS OF NUMBER FIELDS

The following theorem, due to Ostrowski, shows that the valuations of the field of rationals known to us, namely the p-adic valuations and that according to the absolute value, are essentially the only ones possible. Again we assume the values to be real numbers.

A non-trivial valuation φ of the field Γ of rational numbers is either $\varphi(a) = |a|^{\varrho}$ where $0 < \varrho \leqq 1$, i.e., it is equivalent to the ordinary valuation according to absolute values, or it is $\varphi(a) = \varphi_p(a)^{\sigma}$ with a fixed prime p and a fixed positive number σ, and so equivalent to a p-adic valuation.

PROOF. For every rational integer n

$$\varphi(n) \leqq |n|$$

is valid; for we have

$$\varphi(n) = \varphi(|n|) = \varphi(1 + 1 + \cdots + 1) \leqq \varphi(1) + \varphi(1) + \cdots + \varphi(1) = |n|.$$

Let $a > 1$ and $b > 1$ be any two rational integers. We expand b^{ν} according to powers of a:

$$b^{\nu} = c_0 + c_1 a + \cdots + c_n a^n,$$
$$0 \leqq c_{\nu} < a, \quad c_n \neq 0.$$

The highest power a^n of a which is involved is at most equal to b^{ν}:

$$a^n \leqq b^{\nu},$$

i.e.

$$n \leqq \nu \frac{\log b}{\log a}.$$

Now, since

$$\varphi(b^{\nu}) \leqq \varphi(c_0) + \varphi(c_1)\varphi(a) + \cdots + \varphi(c_n)\varphi(a)^n$$
$$< a(1 + \varphi(a) + \cdots + \varphi(a)^n) \leqq a(n+1)M^n,$$

if we take $M = \max(1, \varphi(a))$, we have

$$\varphi(b)^{\nu} < a\left(\frac{\log b}{\log a}\nu + 1\right) M^{\frac{\log b}{\log a}\nu}$$

or

$$\left(\frac{\varphi(b)}{M^{\frac{\log b}{\log a}}}\right)^{\nu} < a\frac{\log b}{\log a}\nu + a.$$

From this follows, by the lemma of Section 73,

$$\varphi(b) \leqq M^{\frac{\log b}{\log a}},$$

i.e.,

$$\varphi(b) \leqq \max\left(1, \varphi(a)^{\frac{\log b}{\log a}}\right).$$

First case. φ is Archimedean. In this case there exists an integer b for which $\varphi(b) > 1$. If, for any other integer $a > 1$, we would have $\varphi(a) \leqq 1$, the contradiction $\varphi(b) \leqq 1$ would follow from the inequality just proved. Therefore, $\varphi(a) > 1$ for all integers $a > 1$. Thus, the inequality in this case is given by

$$\varphi(b) \leqq \varphi(a)^{\frac{\log b}{\log a}}$$

or

$$\varphi(b)^{\frac{1}{\log b}} \leqq \varphi(a)^{\frac{1}{\log a}}.$$

However, since a and b can be interchanged, we also have

$$\varphi(a)^{\frac{1}{\log a}} \leqq \varphi(b)^{\frac{1}{\log b}},$$

and so

$$\varphi(a)^{\frac{1}{\log a}} = \varphi(b)^{\frac{1}{\log b}}.$$

If $\varphi(b) = b^{\varrho}$, it follows that $\varphi(a) = a^{\varrho}$, and therefore

$$\varphi(r) = |r|^{\varrho}$$

for every rational number r. We have $\varrho > 0$ since $\varphi(a) > 1$, and we have $\varrho \leqq 1$, since

$$2^{\varrho} = \varphi(2) = \varphi(1+1) \leqq \varphi(1) + \varphi(1) = 2.$$

Second case. φ *is non-Archimedean.* In this case we have $\varphi(a) \leqq 1$ for all integers a. The totality of all integers a for which $\varphi(a) < 1$ is a prime ideal in the ring of integers. This can be shown exactly as in Section 73: From $\varphi(a) < 1$ and $\varphi(b) < 1$ follows $\varphi(a+b) \leqq \max(\varphi(a), \varphi(b)) < 1$; hence the set of all integral a for which $\varphi(a) < 1$ is an ideal; the ideal is prime, since $\varphi(ab) = \varphi(a)\varphi(b) < 1$ implies either $\varphi(a) < 1$ or $\varphi(b) < 1$ (or both). Now, in the ring of integers every ideal is a principal ideal: in particular, every prime ideal is generated by a prime number. Hence the integers a for which $\varphi(a) < 1$ are exactly the multiples of a prime p. Every rational number r can be written in the form $r = \frac{z}{n} p^{\varrho}$, where z and n are integers not divisible by p. Since $\varphi(z) = \varphi(n) = 1$, we have $\varphi(r) = \varphi(p)^{\varrho} = p^{-\varrho\sigma} = \varphi_p(r)^{\sigma}$, where $\sigma = -\frac{\log \varphi(p)}{\log p}$ is a fixed number, which is positive, since $\varphi(p) < 1$.

For more thorough investigations on valuation theory, only non-Archimedean valuations are of importance. This is seen from the following theorem, due to Ostrowski: *Every field* K *with Archimedean valuation is topologically isomorphic with an ordinary absolute valued field of complex numbers.* For the proof we refer the reader to the original paper.[4]

We shall here deal only with the most important special cases. First of all we consider the field of complex numbers itself.

The field $\Omega = \mathsf{P}(i)$ *of complex numbers has exactly one valuation* Φ *which, for the field* P *of real numbers, coincides with the absolute* valuations $\varphi(a) = |a|^{\varrho}$; this valuation is $\Phi(a) = |a|^{\varrho}$.

---

[4] Ostrowski, A.: "Über einige Lösungen der Funktionalgleichung $\varphi(x)\varphi(y) = \varphi(xy)$." *Acta. math. Vol. 41* (1918) pp. 271-284.

PROOF. Suppose there exists a complex number ξ for which $\Phi(\xi) \neq |\xi|^{\varrho}$.

According as $\Phi(\xi) > |\xi|^{\varrho}$ or $\Phi(\xi) < |\xi|^{\varrho}$, we take $\eta = \frac{\xi}{|\xi|}$ or $\eta = \frac{|\xi|}{\xi}$. Then $|\eta| = 1$ and $\Phi(\eta) > 1$. If we put $\eta^{\nu} = a_{\nu} + ib_{\nu}$, it follows that

$$a_{\nu}^2 + b_{\nu}^2 = |\eta^{\nu}|^2 = |\eta|^{2\nu} = 1,$$
$$|a_{\nu}| \leqq 1, \quad |b_{\nu}| \leqq 1,$$
$$\Phi(a_{\nu}) = |a_{\nu}|^{\varrho} \leqq 1, \quad \Phi(b_{\nu}) = |b_{\nu}|^{\varrho} \leqq 1,$$
$$\Phi(\eta)^{\nu} = \Phi(\eta^{\nu}) \leqq \Phi(a_{\nu}) + \Phi(i)\Phi(b_{\nu}) \leqq 1 + \Phi(i)$$

for all ν. But this is impossible since $\Phi(\eta) > 1$.

We can now construct all Archimedean valuations of an algebraic number field.

All Archimedean valuations of an algebraic number field Σ are obtained by imbedding Σ in all possible ways in the field of real numbers or in that of complex numbers, and by putting $\varphi(a) = |a|^{\varrho}$, $0 < \varrho \leqq 1$ every time.

PROOF. Let Σ be generated by the adjunction of a number ϑ to the field of rational numbers Γ.

By the first theorem of this section, we must have $\varphi(a) = |a|^{\varrho}$ for all elements of Γ, since there are no other Archimedean valuations of Γ. The perfect extension Ω of $\Gamma(\vartheta)$ comprehends the perfect extension of Γ, i.e., the field of real *numbers* P. By the fundamental theorem of algebra, the polynomial $f(x)$ can be decomposed in $\mathsf{P}[x]$ into r_1 linear and r_2 quadratic factors:

$$f(x) = (x - \vartheta_1) \ldots\ldots (x - \vartheta_{r_1})\, q_1(x) \ldots\ldots q_{r_2}(x),$$
$$q_{\nu}(x) = (x - a_{\nu})^2 + b_{\nu}^2.$$

As ϑ is a root of $f(x)$, one of the factors of $f(x)$ must have the root ϑ. If it is one of the linear factors, we have $\vartheta = \vartheta_{\mu}\ (\mu \leqq r_1)$, and $\Gamma(\vartheta)$ appears to be imbedded in the field of real numbers with the valuation $\varphi(a) = |a|^{\varrho}$. However, if ϑ is a root of a quadratic factor $q_{\nu}(x)$, the adjunction of ϑ is equivalent to the adjunction of $i = \sqrt{-1}$, and $\Gamma(\vartheta) = \Sigma$ appears to be imbedded in the field of complex numbers with the valuation $\varphi(a) = |a|^{\varrho}$; for, according to the foregoing, this is the only valuation of $\mathsf{P}(i)$ that continues the absolute valuation $|a|^{\varrho}$ of P.

If ϑ is a root of $q_{\nu}(x)$, ϑ can be imbedded in $\mathsf{P}(i)$ in two different ways, namely by $\vartheta = a_{\nu} + ib_{\nu}$ and by $\vartheta = a_{\nu} - ib_{\nu}$. However, both imbeddings lead to the same valuation, since two conjugate complex quantities have the same absolute value. Accordingly, every linear or quadratic factor of $f(x)$ leads to one, and only one, valuation of the field $\Gamma(\vartheta)$, apart from equivalent valuations. Conversely, every valuation of $\Gamma(\vartheta)$ leads to a definite perfect extension Ω, in which the element ϑ can be a root of only one single factor $x - \vartheta_{\mu}$ or $q_{\nu}(x)$. Thus, the number of the essentially different Archimedean valuations of $\Gamma(\vartheta)$ is equal to the number $r_1 + r_2$ of the factors of $f(x)$. Here r_1 is the number of real roots, while r_2 is the number of pairs of conjugate complex roots of $f(x)$.

The valuations of the field $\Gamma(\vartheta)$ by absolute values are very closely connected with the "units" of this field.[5]

76. VALUATION OF ALGEBRAIC EXTENSION FIELDS

We wish to investigate whether and in how many ways a given non-Archimedean valuation of a field K can be extended to an algebraic extension field Λ of K. Here, as in Section 73, we pass to the exponential valuation $w(a) = \log\varphi(a)$, and by $\mathfrak{p}$ we denote the ideal of the elements a for which $w(a) > 0$ in the ring of the integral elements a for which $w(a) \geqq 0$.

A reducibility criterion, due to Hensel, is of fundamental importance for this investigation. If, in a field with an exponential valuation, a_ν is the coefficient with the smallest exponent of the polynomial

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0,$$

then

$$\frac{a_n}{a_\nu} x^n + \frac{a_{n-1}}{a_\nu} x^{n-1} + \cdots + \frac{a_0}{a_\nu}$$

is a polynomial with integer coefficients, not all of which are divisible by $\mathfrak{p}$. Such a polynomial is called *primitive.*

Reducibility criterion. Let K *be perfect for the exponential valuation* w. *Let* $f(x)$ *be a primitive polynomial with integral coefficients in* K. *Let* $g_0(x)$ *and* $h_0(x)$ *be two polynomials with integral coefficients in* K *which satisfy*

$$f(x) \equiv g_0(x)\, h_0(x) \pmod{\mathfrak{p}}.$$

Then there exist two polynomials $g(x)$, $h(x)$ *with integral coefficients in* K *for which*

$$f(x) = g(x)\, h(x),$$
$$g(x) \equiv g_0(x) \pmod{\mathfrak{p}}$$
$$h(x) \equiv h_0(x) \pmod{\mathfrak{p}},$$

provided $g_0(x)$ *and* $h_0(x)$ *are relatively prime modulo* $\mathfrak{p}$. *It is, moreover, possible to determine* $g(x)$ *and* $h(x)$ *so that the degree of* $g(x)$ *is equal to the degree of* $g_0(x)$ *modulo* $\mathfrak{p}$.

PROOF. Since, without changing hypothesis and conclusion, we may omit in $g_0(x)$ and $h_0(x)$ coefficients divisible by $\mathfrak{p}$, we may assume that $g_0(x)$ is a polynomial of degree r, and that the leading coefficients of $g_0(x)$ and $h_0(x)$ are units. Since again it does not matter whether we replace $g_0(x)$ by $\frac{1}{a} g_0(x)$ and $h_0(x)$ by $a h_0(x)$, we may assume from the beginning that $g_0(x) = x^r + \cdots$ is a *normed* polynomial of degree r. Now, if b is the leading coefficient and s the degree of $h_0(x)$, then the leading coefficient of the product $g_0(x)\, h_0(x)$ is equal to b, and the degree

[5] Cf. B. L. van der Waerden: "Ein logarithmenfreier Beweis des Dirichletschen Einheitensatzes." *Abh. Math. Sem. Hamburg Univ. Vol. 6 (1928)* pp. 259-262.

$r+s \leqq n$. We shall now construct the factors $g(x)$ and $h(x)$ so that $g(x)$ becomes a normed polynomial of degree r, and $h(x)$, therefore, a polynomial of degree $n-r$.

By hypothesis, all the coefficients of the polynomial $f(x)-g_0(x)h_0(x)$ have positive exponents; let the smallest of them be $\delta_1>0$. If $\delta_1=\infty$, then $f(x)=g_0(x)h_0(x)$ so that nothing else need be proved.

Since $g_0(x)$ and $h_0(x)$ are relatively prime modulo $\mathfrak{p}$, there exist two polynomials $l(x)$ and $m(x)$ with integral coefficients in K for which

$$l(x)\,g_0(x)+m(x)\,h_0(x)\equiv 1 \pmod{\mathfrak{p}}$$

holds. Let the smallest of the exponents of the polynomial

$$l(x)\,g_0(x)+m(x)\,h_0(x)-1$$

be $\delta_2>0$. Let the least of the two numbers δ_1, δ_2 be ε, and let, finally, π be an element for which $w(\pi)=\varepsilon$. Then we have

(1) $$f(x)\equiv g_0(x)\,h_0(x) \pmod{\pi}$$

(2) $$l(x)\,g_0(x)+m(x)\,h_0(x)\equiv 1 \pmod{\pi}.$$

We now construct $g(x)$ as the limit of a sequence of polynomials $g_\nu(x)$ of degree r, beginning with $g_0(x)$ and, similarly, $h(x)$ as the limit of a sequence of polynomials $h_\nu(x)$ of degrees $\leqq n-r$, beginning with $h_0(x)$. Suppose that $g_\nu(x)$ and $h_\nu(x)$ have already been so determined that

(3) $$f(x)\equiv g_\nu(x)\,h_\nu(x) \pmod{\pi^{\nu+1}}$$

(4) $$g_\nu(x)\equiv g_0(x) \pmod{\pi}$$

(5) $$h_\nu(x)\equiv h_0(x) \pmod{\pi}$$

and that, furthermore, $g_\nu(x)=x^r+\cdots$ has the highest coefficient 1. For determining $g_{\nu+1}(x)$ and $h_{\nu+1}(x)$, we put

(6) $$g_{\nu+1}(x)=g_\nu(x)+\pi^{\nu+1}u(x)$$

(7) $$h_{\nu+1}(x)=h_\nu(x)+\pi^{\nu+1}v(x).$$

Then we obtain

$$\left|\begin{array}{l} g_{\nu+1}(x)\,h_{\nu+1}(x)-f(x)=g_\nu(x)\,h_\nu(x)-f(x)+ \\ \qquad \pi^{\nu+1}\{g_\nu(x)\,v(x)+h_\nu(x)\,u(x)\}+\pi^{2\nu+2}u(x)\,v(x). \end{array}\right.$$

If, by (3), we put

$$f(x)-g_\nu(x)\,h_\nu(x)=\pi^{\nu+1}\rho(x),$$

we obtain

$$\left|\begin{array}{l} g_{\nu+1}(x)\,h_{\nu+1}(x)-f(x)\equiv \\ \qquad \pi^{\nu+1}\{g_\nu(x)\,v(x)+h_\nu(x)\,u(x)-p(x)\} \pmod{\pi^{\nu+2}}. \end{array}\right.$$

For the left-hand side to become divisible by $\pi^{\nu+2}$ it suffices that the congruence

(8) $$g_\nu(x)\,v(x)+h_\nu(x)\,u(x)\equiv p(x) \pmod{\pi}$$

is satisfied.

In order to achieve this, we multiply congruence (2) by $p(x)$:

(9) $$p(x)\,l(x)\,g_0(x)+p(x)\,m(x)\,h_0(x)\equiv p(x) \pmod{\pi},$$

divide $p(x)m(x)$ by $g_0(x)$ so that the remainder $u(x)$ is of degree $<r$:

(10) $$p(x)\,m(x)=q(x)\,g_0(x)+u(x);$$

next, we substitute (10) in (9),

$$\{p(x)\,l(x) + q(x)\,h_0(x)\}\,g_0(x) + u(x)\,h_0(x) \equiv p(x) \pmod{\pi},$$

replace by zero all coefficients of the polynomial in braces which are divisible by π and obtain

$$v(x)\,g_0(x) + u(x)\,h_0(x) \equiv p(x) \pmod{\pi}. \tag{11}$$

From (11) follows the desired congruence (8) because of (4) and (5). Furthermore, $u(x)$ is of degree $< r$, and so, because of (6), $g_{r+1}(x)$ is of the same degree and has the same leading term as $g_r(x)$. It remains to be shown that $v(x)$ is of degree $\leqq n - r$. If this were not the case, a highest term of degree $> n$ would occur in the first term of (11) but not in the others. By (11), the coefficient of this term would have to be divisible by π so that the leading coefficient of $v(x)$ would be divisible by π. But since we omitted in $v(x)$ all coefficients divisible by π, $v(x)$ is of degree $\leqq n - r$.

From the congruence (8) follows, as we saw before,

$$f(x) \equiv g_{r+1}(x)\,h_{r+1}(x) \pmod{\pi^{r+2}}. \tag{12}$$

Because of (3), (6), (7) the sequences $\{g_r(x)\}$ and $\{h_r(x)\}$ thus constructed converge to polynomials $g(x) = x^r + \cdots$ and $h(x)$ so that

$$f(x) = g(x)\,h(x).$$

Furthermore, because of (4) and (5), we have

$$g(x) \equiv g_0(x) \pmod{\mathfrak{p}}$$
$$h(x) \equiv h_0(x) \pmod{\mathfrak{p}}.$$

A simple deduction from the reducibility criterion is the following:

For an irreducible polynomial

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n$$

in K *we have*

$$\min\,(w(a_0), w(a_1), \ldots, w(a_n)) = \min\,(w(a_0), w(a_n)).$$

To prove this, we may assume that $f(x)$ is primitive. Then the $\min(w(a_0), \ldots, w(a_n))$ is zero. Suppose that $w(a_0)$ and $w(a_n)$ were both greater than zero; then an r, $0 < r < n$, would exist for which $w(a_r) = 0$, but $w(a_\nu) > 0$ for $\nu = r+1, \ldots, n$. Then we would have

$$f(x) \equiv (a_0 + a_1 x + \cdots + a_r x^r) \cdot 1 \pmod{\mathfrak{p}}$$
$$0 < r < n,$$

and therefore, by the reducibility criterion, $f(x)$ would be decomposable into a factor of degree r and one of degree $n - r$.

EXERCISES. 1. If a polynomial $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ has integral coefficients in K and is irreducible mod $\mathfrak{p}$, then $f(x)$ is irreducible in the complete field Ω_{K} as well.

2. If all coefficients $a_{n-1}, \ldots, a_0$ in $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ are divisible by $\mathfrak{p}$, and if a_0 is not the product of two elements in $\mathfrak{p}$, then $f(x)$ is irreducible (generalization of Eisenstein's irreducibility criterion).

3. Examine the decomposition of the rationally irreducible polynomials

$$x^2+1,\quad x^2+2,\quad x^2-3,\quad x^2+x+1,\quad x^3+2$$

in the field of triadic numbers.

The most important application of the foregoing theorem is the extension of complete exponential valuations to algebraic extensions.

Let K *be complete relative to the exponential valuation* w, *and let* Λ *be an algebraic extension of* K. *Then there exists an exponential valuation* W *of* Λ *which coincides with* w *within* K .

PROOF. 1. Let ξ be an element of Λ, and let

$$\xi^n+a_{n-1}\xi^{n-1}+\cdots+a_0=0$$

be the irreducible equation for ξ with coefficients in K. We assert that

$$W(\xi)=\frac{1}{n}\,w(a_0)$$

is a valuation of Λ (which is obviously identical with w relative to K). In order to prove the relations

$$W(\xi\eta)=W(\xi)+W(\eta)$$
$$W(\xi+\eta)\geqq\min(W(\xi),\ W(\eta))$$

for two elements ξ, η of Λ , we consider the subfield $\Lambda_0=\mathsf{K}(\xi,\eta)$, which is of finite degree t over K . By Section 41, $W(\xi)$ can also be defined thus:

$$W(\xi)=\frac{1}{t}w(N_{\Lambda_0/\mathsf{K}}(\xi)).$$

Thence it follows at once that

$$W(\xi\eta)=W(\xi)+W(\eta),$$

since $N(\xi\eta)=N(\xi)N(\eta)$. In the proof of $W(\xi+\eta)\geqq\min(W(\xi),\ W(\eta))$ we may limit ourselves to $\eta=1$ since

$$W(\xi+\eta)=W(\eta)+W\left(1+\frac{\xi}{\eta}\right)$$

and since $\min(W(\xi),\ W(\eta))=W(\eta)+\min\left(W\left(\frac{\xi}{\eta}\right),\ 0\right)$.

Now the irreducible equation for $\xi+1$ is given by

$$(\xi+1)^n+\cdots+(a_0-a_1+a_2-\cdots+(-1)^{n-1}a_{n-1}+(-1)^n)=0.$$

By the preceding theorem we have indeed

$$W(\xi+1)=\frac{1}{n}w(a_0-a_1+\cdots)\geqq\frac{1}{n}\min(w(a_0),w(a_1),\ldots,w(a_{n-1}),w(1))$$
$$=\frac{1}{n}\min(w(a_0),\ w(1))=\min(W(\xi),\ 0).$$

Returning from the exponential valuations $w(a)$, $W(\xi)$ to the ordinary valuations

$$\varphi(a)=e^{w(a)},\quad \Phi(\xi)=e^{W(\xi)},$$

we see that the valuation of the extension field Λ is defined by

$$\Phi(\xi)=\sqrt[n]{\varphi(a_0)}$$

or, if Λ is of finite degree t over K, by

$$\Phi(\xi)=\sqrt[t]{\varphi(N_\Lambda(\xi))}.$$

We note that the very same formula is also correct for Archimedean valuations if K is the field of real numbers, Λ the field of complex numbers, and if $\Phi(\xi)=|\xi|^e$. For in this case we have

$$|\xi|=\sqrt{a^2+b^2}=\sqrt{N(\xi)}=\sqrt{|N(\xi)|}$$

for $\xi=a+bi$.

Therefore, we shall from now on treat Archimedean and non-Archimedean valuations side by side; thus, by K we shall mean either a complete non-Archimedean field with a valuation, or the field of real numbers, or that of complex numbers with absolute valuations $\varphi(a)=|a|^e$.

Let Λ be of finite degree over K, and let $u_1,\dots,u_n$ be a basis of Λ/K. Let K be complete for the valuation φ. If Φ is a valuation of Λ which coincides with φ within K, then a sequence

$$c_\nu=a_1^{(\nu)}u_1+\cdots+a_n^{(\nu)}u_n,\quad \nu=1,2,3,\dots$$

is a fundamental sequence for Φ if, and only if, the n sequences $\{a_i^{(\nu)}\}$ are fundamental sequences for φ.

Since the sequences $\{a_i^{(\nu)}\}$ converge to limits a_i in K, it follows that Λ *is complete for Φ*.

PROOF. We prove the convergence of the sequences $\{a_i^{(\nu)}\}$ by complete induction. If the c_ν are of the form

$$c_\nu=a_1^{(\nu)}u_1,$$

then $\{a_1^{(\nu)}\}$ is of course a fundamental sequence as soon as $\{c_\nu\}$ is one. Suppose the assertion is true for all sequences $\{c_\nu\}$ of the form

$$c_\nu=\sum_{i=1}^{m-1}a_i^{(\nu)}u_i.$$

Let a sequence

$$c_\nu=\sum_{i=1}^{m}a_i^{(\nu)}u_i$$

be given. If the sequence $\{a_m^{(\nu)}\}$ converges, then $\{c_\nu-a_m^{(\nu)}u_m\}$ is a fundamental sequence as well and thus the $\{a_i^{(\nu)}\}$, where $i<m$, converge by the induction assumption. Suppose $\{a_m^{(\nu)}\}$ were not convergent. In this case it would be possible to choose the number sequence $n_1, n_2, n_3,\dots$ so that $\varphi(a_m^{(\nu)}-a_m^{(\nu+n_\nu)})>\varepsilon$ for all ν, where ε is a fixed *positive number*. Therefore, the sequence

$$d_\nu=\frac{c_\nu-c_{\nu+n_\nu}}{a_m^{(\nu)}-a_m^{(\nu+n_\nu)}}=\sum_{i=1}^{m-1}\frac{a_i^{(\nu)}-a_i^{(\nu+n_\nu)}}{a_m^{(\nu)}-a_m^{(\nu+n_\nu)}}u_i+u_m=\sum_{i=1}^{m-1}b_i^{(\nu)}u_i+u_m$$

would have to converge to zero; for the sequence of the numerators converges to zero, since $\{c_\nu\}$ is a fundamental sequence. Now we have

$$d_\nu - u_m = \sum_{i=1}^{m-1} b_i^{(\nu)} u^i .$$

Thus, by the induction assumption, the sequences $\{b_i^{(\nu)}\}$ converge to certain limits b_i, and we would have

$$-u_m = \sum_{i=1}^{m-1} b_i u_i ,$$

contrary to the fact that $u_1, \ldots, u_n$ is a basis of Λ/K.

In the very same manner we prove the following: The sequence $\{c_\nu\}$ is a null sequence if, and only if, the sequences $\{a_i^{(\nu)}\}(i = 1, \ldots, n)$ are null sequences.

On this assertion rests the proof of the following uniqueness theorem:

The continuation Φ of the valuation φ of a perfect field K to an algebraic extension Λ is uniquely determined, namely by

$$\Phi(\xi) = \sqrt[n]{\varphi(N(\xi))},$$

where the norm is to be formed in the field $\mathsf{K}(\xi)$, and where n is the degree of this field over K.

PROOF. It will suffice to consider a fixed element ξ with the corresponding field $\mathsf{K}(\xi)$; by norms we shall always means norms in this field. If a sequence $\{c_\nu\}$ in this field tends towards zero (in the sense of Φ), and if we express the c_ν linearly in terms of the basis elements $u_1, \ldots, u_n$ of $\mathsf{K}(\xi)$, then, by the above, the single coefficients $a_i^{(\nu)}$ tend toward zero, and so does the norm, being a homogeneous polynomial in these coefficients. Suppose now we had $\Phi(\xi)^n < \varphi(N(\xi))$ or $\Phi(\xi)^n > \varphi(N(\xi))$. By considering the elements

$$\eta = \frac{\xi^n}{N(\xi)}, \quad \text{or} \quad \eta = \frac{N(\xi)}{\xi^n},$$

respectively, we see that in either case $N(\eta) = 1$ and $\Phi(\eta) < 1$. It follows that $\lim \eta^\nu = 0$ so that $\lim N(\eta^\nu) = 0$, giving a contradiction to $N(\eta^\nu) = N(\eta)^\nu = 1$.

From the theorem just proved follows:

An isomorphism between two valued algebraic extension fields Λ, Λ' of the perfect valued field which carries the elements of K into themselves necessarily carries the valution of Λ into that of Λ'.

In Section 75 we constructed the Archimedean valuations of an algebraic number field by imbedding the field into a complete field which could then be identified, respectively, with the field of real numbers, or with that of complex numbers. In an entirely similar manner we are going to study the continuations of a non-Archimedean valuation of an incomplete field K to an algebraic extension field by extending the former to a complete field. For simplicity, we limit ourselves to the case of a finite extension Λ of the field K, which may arise by the adjunction of $\xi_1, \ldots, \xi_t$ to

K, although all the following considerations may readily be applied to the case, in which a countable (or well-ordered) sequence of quantities $\xi_1, \xi_2, \ldots$ is adjoined to K.

Let $\xi_1, \ldots, \xi_t$ be roots of the polynomials $f_1(x), \ldots, f_t(x)$. Let the complete extension field belonging to K be Ω. Let the decomposition field of $f_1(x), \ldots, f_t(x)$ over Ω be Σ. Σ, as an algebraic extension of the complete field Ω has a single valuation, which continues the valuation of Ω. The same is true for all intermediate fields between Ω and Σ; the valuation of such an intermediate field is given by that of Σ.

Clearly, we can obtain a valuation of $\Lambda = \mathsf{K}(\xi_1, \ldots, \xi_t)$, as soon as we succeed in *imbedding* Λ in Σ, i.e., in finding an isomorphism σ of Λ which carries Λ into a subfield $\Lambda' = \mathsf{K}(\xi_1', \ldots, \xi_t')$ of Σ and, at the same time, leaves fixed the elements of K. Then, by virtue of σ^{-1}, the valuation of Λ' as subfield of Σ is immediately transferred to Λ.

We now assert that all valuations of Λ continuing that of K are already exhausted by these imbeddings.

To every such valuation φ of Λ there exists an isomorphism σ which carries Λ into a subfield Λ' of Σ, leaves the elements of K fixed, and carries the valuation φ of Λ into the valuation of Λ'.

PROOF. We form the perfect extension of Λ. It includes the perfect extension Ω of K and contains the quantities $\xi_1, \ldots, \xi_t$; therefore, it includes the field $\Omega(\xi_1, \ldots, \xi_t)$. Now, this field can always be extended to a decomposition field of the polynomials $f_1, \ldots, f_t$ which is isomorphic with the decomposition field Σ. The isomorphism carries $\Omega(\xi_1, \ldots, \xi_t)$ into a subfield $\Omega(\xi_1', \ldots, \xi_t')$, leaves all elements of Ω fixed, and therefore carries the valuation of $\Omega(\xi_1, \ldots, \xi_t)$ into the only possible valuation of $\Omega(\xi_1', \ldots, \xi_t')$.

From the theorem just proved it follows that *all possible continuations of the given valuation of K to valuations of $\Lambda = \mathsf{K}(\xi_1, \ldots, \xi_n)$ can be found by imbedding the field Λ in the complete decomposition field Σ in all possible ways.*

If, in particular, Λ is a simple extension $\Lambda = \mathsf{K}(\xi)$ with the defining equation $f(\xi) = 0$, Σ is the decomposition field of the polynomial $f(x)$. Let the decomposition of $f(x)$ into irreducible factors in Ω be given by

$$f(x) = f_1(x) f_2(x) \ldots f_s(x).$$

Every isomorphism σ of $\mathsf{K}(\xi)$ carries the element ξ into a root of the polynomial $f(x)$, i.e., into a root of one of the polynomials $f_\nu(x)$. To each of the polynomials $f_\nu(x)$ belongs an extension field $\Omega(\xi')$ which has a uniquely determined valuation; here ξ' may be any root of $f_\nu(x)$ whatsoever. Every isomorphism $\mathsf{K}(\xi) \cong \mathsf{K}(\xi')$ which carries ξ into ξ' and leaves fixed the elements of K defines an imbedding in the above sense and, therefore, a possible valuation of $\mathsf{K}(\xi)$. Accordingly, there are (exactly as in Section 75) just as many different valuations as there are irreducible factors of $f(x)$ in $\Omega[x]$.

The case of a non-simple algebraic extension can be reduced to successive simple extensions by adjoining $\xi_1, \ldots, \xi_t$ in serial order. We may equally well decompose a non-simple extension into a separable extension which, by the theorem on the primitive element, is simple and into a subsequent adjunction of p-th roots ($p =$ characteristic of K) to which the valuation is uniquely conveyed in a trivial fashion, since then *every* element of the extension field is a p-th root, and since the valuation is given by

$$\varphi\left(\sqrt[p]{a}\right) = \sqrt[p]{\varphi(a)}.$$

Recent Publications on Fields With Valuations.

Ostrowski, A.: "Untersuchungen zur arithmetischen Theorie der Körper. (Die Theorie der Teilbarkeit in allgemeinen Körpern)." *Math. Z. Vol. 39 (1934) pp. 269-404.*

Witt, E.: "Zyklische Körper und Algebren der Charakteristik p vom Grad p^x." *J. reine angew. Math. Vol. 176 (1936) pp. 126-140.*

Teichmüller, O.: "Diskret bewertete perfekte Körper mit unvollkommenem Restklassenkörper." *J. reine angew. Math. Vol. 176 (1936) pp. 141-152.*

Mahler, K.: "Über Pseudobewertungen." *I. Acta math. Vol. 66 (1936) pp. 79-199; Ia. Akad. Wetensch. Amsterdam. Proc. Vol. 39 (1936) pp. 57-65; II. Acta math. Vol. 67 (1936) pp. 51-80.*

Krull, W.: "Allgemeine Bewertungstheorie." *J. reine angew. Math. Vol. 167 (1932) pp. 160-196.*

77. VALUATIONS OF ALGEBRAIC NUMBER FIELDS

As an application of the general theory of the preceding section, all possible valuations—Archimedean and non-Archimedean—of any algebraic number field, i.e., of any finite extension of the rational number field will be determined.

Let Γ be the field of rational numbers, and Λ an algebraic number field

$$\Lambda = \Gamma(\vartheta).$$

Let the generating element ϑ be a root of an irreducible polynomial $f(x)$.

First, we determine, according to Section 75, all possible valuations of Γ. Apart from equivalent valuations, we have only one Archimedean valuation, viz.

$$\varphi(a) = |a|,$$

and for every prime p one p-adic valuation

$$\varphi(a) = (\varphi_p(a).$$

$\varphi_p(a)$ is p^{-n} if p^n is the highest power of p which divides the number a.

Now the question is, how these valuations can be extended to Λ.

For the Archimedean valuation this question was solved at the end of Section 75. The complete Archimedean extension of Γ is the field of real numbers P. In this field $f(x)$ may be decomposed into linear and quadratic factors:

$$f(x) = (x-\vartheta_1)\ldots\ldots(x-\vartheta_r)\,q_1(x)\ldots\ldots q_s(x),$$
$$q_\nu(x) = (x-a_\nu)^2 + b_\nu^2.$$

We now find all Archimedean valuations of $\Lambda = \mathsf{P}(\vartheta)$ by imbedding Λ in all possible ways in the real or complex field P or $\mathsf{P}(i)$, i.e., by identifying ϑ either with a real ϑ_μ $(\mu = 1, 2, \ldots, r)$, or with a complex root

$$\vartheta_{r+\nu} = a_\nu \pm i b_\nu \qquad (\nu = 1, 2, \ldots s).$$

The sign $+$ or $-$ does not matter, since complex conjugates have the same absolute value. In each case every element $\alpha = g(\vartheta)$ is identified with a real or complex number $\alpha_\mu = g(\vartheta_\mu)$ $(\mu = 1, \ldots, r+s)$, and the value of α in the resulting valuation is the absolute value of α_μ:

$$\varphi_\mu(\alpha) = |\alpha_\mu| \qquad (\mu = 1, 2, \ldots, r+s).$$

In algebraic number theory the complex numbers α_μ are called the *real and complex conjugates* of the algebraic number α. Thus, the valuations are simply the absolute values of the real and complex conjugates of α.

The non-Archimedean valuations can be treated in exactly the same way. To the p-adic valuation of Γ we first construct the complete extension of Ω_p — the field of p-adic numbers. In this field the polynomial $f(x)$ can be decomposed into irreducible factors:

$$f(x) = f_1(x) f_2(x) \ldots f_s(x). \tag{1}$$

We choose arbitrarily a root ϑ_1 of $f_1(x)$, ϑ_2 of $f_2(x)$, ..., ϑ_s of $f_s(x)$, each in a simple extension field of Ω_p. All valuations of K can now be obtained, as we have seen in Section 76, by identifying ϑ with any one of the ϑ_μ $(\mu = 1, 2, \ldots s)$ i.e., by means of isomorphisms σ_μ which carry ϑ into ϑ_μ. These isomorphisms carry the elements $\alpha = g(\vartheta)$ into $\alpha_\mu = g(\vartheta_\mu)$, and the valuations are defined by

$$\Phi(\alpha) = \Phi(\alpha_\mu),$$

where $\Phi(\alpha_\mu)$ is the n-th root of the norm of α_μ taken in the field $\Omega_p(\alpha_\mu)$ with respect to Ω_p.

The only remaining question is, how the factorization (1) can be found. This is very simple, provided p is not a divisor of the discriminant of the polynomial $f(x)$. We can always suppose the coefficients $a_0, a_1, a_2, \ldots, a_n$ of $f(x)$ to be rational integers. Replacing ϑ by $a_0\vartheta$, we may even assume $a_0 = 1$. Now, if p does not divide the discriminant, $f(x)$ can be decomposed modulo p into relatively prime factors, irreducible modulo p:

$$f(x) \equiv g_1(x)\, g_2(x) \ldots g_s(x) \pmod{p}. \tag{2}$$

According to the reducibility criterion of Section 76, a corresponding factorization of $f(x)$ in the p-adic field Ω_p can be found, viz.

$$f(x) = f_1(x) f_2(x) \ldots f_s(x).$$

in which $f_1(x), \ldots, f_s(x)$ are congruent (modulo p) to $g_1(x), \ldots, g_s(x)$. It is clear that $f_1, \ldots, f_s$ are irreducible in Ω_p, for they are even irreducible modulo p. Thus, we have found the required factorization (1).

Only a finite number of primes p, namely the prime divisors of the discriminant, cause more difficulties. Let the factorization of $f(x)$ modulo p be

$$f(x) \equiv g_1(x)^{e_1} g_2(x)^{e_2} \dots g_t(x)^{e_t} \pmod{p}$$

According to the reducibility criterion of Section 76, we have in this case

$$f(x) = G_1(x) G_2(x) \dots G_t(x),$$

with

$$G_k(x) \equiv g_k(x)^{e_k} \pmod{p}.$$

In many cases the irreducibility of the polynomials G_k can be proved by the aid of the generalized Eisenstein Criterion (Section 78, Exercise 2). In other cases one can try to find factorizations modulo p^2, mod p^3, etc., and thus eventually reach a p-adic factorization, just as in the proof of the Criterion of Reducibility.

As an example, we shall derive all possible valuations of Gauss' number field. This field consists of all complex numbers $a+bi$, a and b being rational. The irreducible polynomial with root i is x^2+1. In this field of real numbers, x^2+1 is irreducible; hence there is only one Archimedean valuation, viz:

$$\varphi(a+bi) = \sqrt{N(a+bi)} = \sqrt{a^2+b^2} = |a+bi|.$$

The discriminant of x^2+1 is 4; hence the only prime number dividing the discriminant is 2. In the field Ω_2 the polynomial

$$(x+1)^2+1 = x^2+2x+2$$

is irreducible according to Eistenstein's Criterion (Section 76, Exercise 2). Hence the 2-adic valuation $\varphi_2(a)$ of the rational number field can be extended to Gauss' field in only one way, viz:

$$\Phi_2(a+bi) = \sqrt{\varphi_2(a^2+b^2)}.$$

This valuation is identical with the $(1+i)$-adic valuation generated by the Gaussian prime number $(1+i)$.

Now let p be an odd prime. The p-adic valuation of the rational number field can be extended to Gauss' field in 1 or 2 ways, according as the polynomial (x^2+1) remains irreducible modulo p or decomposes into 2 linear factors. It decomposes as soon as it has a root $r \pmod{p}$, i.e. as soon as an integer r exists such that

$$r^2+1 \equiv 0 \pmod{p},$$

or

$$r^2 \equiv -1 \pmod{p}.$$

In this case -1 is said to be a *quadratic residue* modulo p. Now the decomposition modulo p is

$$x^2+1 \equiv (x+r)(x-r) \pmod{p}$$

According to the Criterion of Reducibility, this decomposition modulo p implies the existence of a p-adic factorization

$$x^2+1 = (x+\vartheta)(x-\vartheta),$$

in which ϑ is a p-adic power series (Section 74), starting with r as its first term. In this case 2 different valuations are obtained by identifying i either with ϑ or with $-\vartheta$:

$$\Phi_1(a+bi) = \varphi_p(a+b\vartheta)$$
$$\Phi_2(a+bi) = \varphi_p(a-b\vartheta),$$

φ_p being the p-adic valuation of the field Ω_p.

On the other hand, if x^2+1 remains irreducible (mod p), there is only one p-adic valuation:

$$\Phi(a+bi) = \sqrt{\varphi_p(a^2+b^2)}.$$

EXERCISES. 1. According to Section 37, Exercise 3, the group of remainder classes modulo p, excluding zero, is a cyclic group of order $(p-1)$. Show that -1 is the only element of order 2 in this group, and that -1 is a quadratic residue (modulo p) if, and only if, $(p-1)$ is divisible by 4.

2. Find all valuations of the field $\Gamma(\sqrt{2})$.

3. Find all valuations of the field of third roots of unity $\Gamma(\sqrt{-3})$.

INDEX

(The numbers refer to the pages on which the expressions appear for the first time.)

Abelian (commutative), 11
Abelian equation, 155
  extension field, 155
  group, 11
Abel's theorem, 177
Absolute value, 209
  of a complex number, 224
Additive group (module), 15
  of a ring, 33
Adjunction, 94
  of an indeterminate, 47
  symbolic, 99
Admissible normal divisor, 138
Admissible subgroup, 138
Algebra (hypercomplex system), 42
Algebraic basis, 201
  field extension, 106
  function, 204
  number, 222
  number field, 222
  quantity, 96
  with respect to a field, 95
Algebraically closed, 193
  dependent, 200
  equivalent, 201
Algorithm, Euclidean, 57
Alternating group, 19
Archimedean ordering, 210
  valuation, 237
Archimedes' axiom, 210
Arithmetic series of the n-th order, 69
Associate, 59
Associative law
  for groups, 11
  of addition, 4, 32
  of multiplication, 4, 32
Automorphism, 25
  inner, 26
  outer, 26
Automorphism
  group, 25
  ring, 139
Axioms, Peano's, 4

Basis element, 42
Basis, field, 102
  of a vector space, 43
Binomial theorem, 36
Biunique mapping, 3
Bound
  least upper, 211
  lower, 211
  upper, 211
Bounded, 211

Cantor's construction of the real numbers, 212
Cardano's solution formulae, 179
Cardinality (cardinal number), 3
Casus irreducibilis, 180
Cauchy's convergence theorem, 216, 240
Center (central) of a group, 24
Chain rule, 207
Characteristic, 92
Characteristic subgroup, 139
Class, 1
Coefficients, 46
Combination, rules of, 11
Commutative, 33
Commutative law, 4, 32
Commutator subgroup, 30
Complete field extension, 240
Completely reducible group, 148
Complex, 22
Complex numbers, 222
Components (of a vector), 43
Composition factors, 143
Composition series, 143
  isomorphism of a, 143
Congruence
  relative to ideals, 50
  relative to modules, 30
Conjugate
  fields, 97
  group elements, 26
  resolvents, 170
  subgroup, 26
Construction
  by complete induction, 5
  of a polygon of 17 sides, 168
  of regular polygons, 186
Constructions with ruler and compass, 183
Content of a polynomial, 71
Continuation of isomorphisms, 107
Continuous, 218
Coset, 22
Countable (denumerable), 8
Cubic resolvent, 181

Cycle, 19
Cyclic
  equation, 155, 168
  field, 155, 168
  group, 20
Cyclotomic equations
  Galois group of the, 162
  irreducibility of the, 162
  periods of the, 163
Cyclotomic field, 160
Cyclotomic polynomial, 114

Decomposition field, 107
Defining equation, 96
Definition by complete induction, 5
Degree
  formal, 83
  of a finite extension, 103
  of a permutation group, 152
  of a polynomial, 46
  of a rational function, 197
  of an algebraic quantity, 96
  of transcendence, 201
Denumerable, 8
Dependence
  algebraic, 200
  linear, 100
Derivative
  of an algebraic function, 203
  of a polynomial, 63
Difference product, 177
Differences, array of, 68
Differential quotient (derivative), 63, 202
Differentiation, 63
  of algebraic functions, 202
  total, 203
Dimension (of a vector space), 43, 103
Direct product, 146
Discriminant, 82
Disjoint sets, 2
Distributive law, 4, 18, 32
Divisibility
  of elements, 53, 239
  of ideals, 53
Division, 14
Division algebra, 44
Division algorithm, 48
Divisor, 53
  greatest common, 55
  proper, 53, 59
Domain of definition, 3
Domain of integrity, 34
Duplication of the cube, 185

Eisenstein's theorem, 74
  (generalized), 250
Element
  indecomposable, 59
  inverse, 33
  of a set, 1
  prime, 59
Elementary symmetric function, 78
Elimination, 85
  successive, 105
Empty set, 1
Endomorphism, 28
Equation
  Abelian, 155
  cyclic, 155, 168
  generic, 175
  metacyclic, 175
  normal, 111
  of the second, third, fourth degrees, 177
  primitive, 155
  pure, 168
  soluble, 173
Equinumerous sets, 3
Equivalence relation, 10
Equivalent, 97
  extensions, 97
  sets, 101
Euclidean
  algorithm, 57
  ring, 56
Euler's differential equation, 64
Euler's function, 112
Explicitly given field, 134
Exponent, 120, 123
Exponential valuation, 238
Extension
  algebraic, 106
  finite, 102, 106
  infinite, 193
  normal, 109
  simple, 94
  simple algebraic, 95
  simple transcendental, 95
Extension field, 93

Factor group, 29
Factorization, 58
  in a finite number of steps, 77
  of polynomials, 70
Factors of a normal series, 142
Fermat's theorem, 118
Field (domain of rationality), 37
  Abelian, 155
  cyclic, 155, 168
  degree of a, 103
  of algebraic numbers, 222
  of complex numbers, 222
  of p-adic numbers, 241
  of quotients, 39
  of rational numbers, 44, 92
  of rationality, 37
  of real algebraic numbers, 232
  of real numbers, 217
Field basis, 102
Field extension, 93

Fields with a valuation, 235
Finite commutative field, 115
Finite rank, 102
Form (homogeneous polynomial), 48
Formal
  degree, 83
  leading coefficient, 83
Formally real, 225
Four(s) group, 27
Function, 2
  continuous, 218
Fundamental sequence, 212
Fundamental theorem
  of algebra, 223
  of the Galois theory, 156
  on normal series, 144
  on symmetric functions, 79

Galois
  field, 115
  group, 154
  resolvent, 111
  theory, 153
Gauss'
  formula, 166
  theorem, 71
Gaussian
  integers, 44
  number field, 44
  sum, 168
G.C.D. (greatest common divisor), 55, 56
Generic equation of degree n, 175
Generated
  ideal, 49
  subgroup, 20
Greater and less, 5, 209
Group, 11
  Abelian, 11
  additive, 15
  alternating, 19, 149
  completely reducible, 148
  cyclic, 20
  finite, 14
  imprimitive, 150
  intransitive, 150
  of a field, 154
  primitive, 151
  simple, 142
  soluble, 146
  symmetric, 13
  transitive, 150
  with operators, 138
Group ring, 45

Hensel's
  p-adic numbers, 241
  reducibility criterion, 248
Homogeneous polynomial, 48
Homomorphic
  group, 28
  ring, 38
Homomorphism
  of groups, 28
  of rings, 38
Homomorphism, law of
  for groups, 30, 140
  for rings, 53

Ideal, 49
  left, 49
  maximal, 54
  right, 49
  two-sided, 49
Ideal basis, 50
Identity (identical transformation), 13
  of a group, 11
  of a ring, 35
Image, 3
Imperfect, 124
Imprimitive group, 150
Imprimitivity, system of, 151
Indecomposable, 59, 73
Independent transcendental, 202
Indeterminate, 45
Index of a subgroup, 23
Induction, complete, 4
Infinite
  field extension, 193
  set, 7
Infinitely large (small) elements, 210
Inner automorphism, 26
Inseparable (of the second kind), 120
Integers, 6
  p-adic, 242
Integral domain, 34
Intermediate
  field, 103, 156
  group, 27
Interpolation formula, 66
  Lagrange's, 66
  Newton's, 66
Intersection, 2
Intransitive group, 150
Invariant subgroup, 24
Inverse
  element, 11, 37
  image, 3
Irreducibility criterion
  Dumas', 76
  Eisenstein's, 74
Irreducible polynomials, 59
Isobaric, 80
Isomorphic
  groups, 25
  normal series, 143
  rings, 38
  sets, 24

Isomorphism
first law of, 141
many-one, 38
second law of, 142
topological, 244

Jordan-Hölder theorem, 145

Klein's four(s) group, 27

Lagrange resolvent, 169
Lagrange's interpolation formula, 66
Law of homomorphism
for groups, 30, 140
for rings, 53
Law of the mean, 222
L.C.M. (least common multiple), 55
Leading coefficient (highest coefficient), 46
formal, 83
Least upper bound, 211
theorem on the, 216
Left ideal
ideal, 49
inverse, 35
Length of a normal series, 142
Less than, 5, 209
Lexicographically, 81
Limit, 215
Linear
form, 104
rank, 102
substitutions, fractional, 198
substitutions modulo, 9
Linearly
dependent, 100, 101
independent, 100, 101
Lower bound, 211
Lüroth's theorem, 198

Mapping, 3
biunique, 3
inverse, 3
Matric ring, 45
Matrix, 45
Maximal, 54
Metacyclic, 175
equations of prime degree, 187
Minimal basis, 102
Möbius' function, 114
Module (additive group), 15
basis, 102
homomorphism, 140
of linear forms, 42
property, 49
Modulo, 50
Multiple, 35, 53
least common, 55
proper, 53
Multiplication table, 15
Multiplicative group of a skew field, 37
Multiplicity of a zero (*or* root), 109
Multipliers, domain of, 139

Natural number, 3
Negative, 209
number, 6
Newton's interpolation formula, 66
Norm, 128
of a Gaussian number, 58
reduced, 133, 134
Normal divisor, 24
admissible, 137
Normal equation, 111
Normal extension field, 109
Normalizer, 27, 30
Normal series, 142
factor of a, 142
fundamental theorem on, 144
length of a, 142
refinement of a, 143
without repetition, 143
Null element, 33
ideal, 49
sequence, 213
Nullstellensatz, Weierstrass', 218
Number module, 12
Number of elements, 8
Numbers
algebraic, 222
complex, 44, 222
natural, 3
p-adic, 241
rational, 42
real, 9, 218
sequence, 3

Object of a transformation, 12
One-to-one correspondence, 3
Operation, rules of, 32
Operator
domain, 138
homomorphism, 140
isomorphism, 140
Operators, 138
Order-isomorphic, 210
Order
of a group, 15
of an element, 21
Ordered fields, 209

P-adic
integer, 242
number, 241
valuation, 236
Partial fraction decomposition, 88
Partitions, 9
Peano, axioms of, 4
Perfect, 121, 124

Periods of a cyclotomic equation, 163
Permutation, 12
  even, 19
  odd, 19
Polynomial, 45
  irreducible, 59
  primitive, 71
  regular, 73
  separable, 120
Polynomial
  domain, 46
  ring, 46
Positive(ness), 209
Power
  remainders, 118
  series, formal, 243
Powers, 16, 35
  sums of, 81
Pre-image, 3
Prime
  element, 59
  field, 91
  ideal, 53
  number, 59
Primitive
  element, 126
  equation, 155
  group, 151
  polynomial, 73
Principal ideal, 49
  ring, 56
Principal theorem on finite sets, 7
Product, 11
  composite, 15
  direct, 146
  of two complexes, 22
  of two numbers, 4
  of two transformations, 13
Proper divisor, 53, 59
  multiple, 53
  subset, 2
Pure equation, 168
Pure transcendental extension, 202

Quadrature of the circle, 186
Quaternion field, 45
Quaternions, 45
Quotient
  group, 29
  ring, 42

Radical, 172
  irreducible, 174
Range of values of a function. 3
Rank of a set of equations, 104
  linear, 102
Rational curve, 199
  integral function, 47
  numbers, 42
Real closed, 225
Real field, 225
  number, 9, 218
Recursion relations, 5
Reduced degree
  of a field, 123
  of a polynomial, 120
Reduced
  norm, 133
  trace, 133
Reducibility criterion by Hensel, 248
Refinement of a normal series, 143
Reflexive, 9
Relative isomorphism, 121
Relatively prime, 57
Replacement theorem, 101
Representative, 10
Residue class
  module, 30
  ring, 52
  sequence, 241
Residue classes
  for groups, 30
  for rings, 50
Resolvent
  cubic, 181
  Lagrange, 169
Resultant, 84
  indecomposability of the, 87
Right
  ideal, 49
  inverse, 35
  multiple, 49
Ring, 32
  adjunction, 47
  homomorphism, 38
  with identity, 35
  without divisors of zero, 34
Rolle's theorem, 222
Root
  field, 125
  of a polynomial, 64, 96
  of multiplicity k, 65
Root of unity, 111
  primitive, 112
Rules of operation, 32

Schreier's fundamental theorem on normal series, 144
Segment of the number sequence, 6
Separable
  extension, 120
  of the first kind, 120
  polynomial, 120
Set, 1
  countable (denumerable), 8
  finite, 7
  infinite, 7

Sign
of a number, 220
variation in, 220
Similarly ordered, 25
Simple
field extension, 94
group, 142
transcendental extension, 95, 197
Simplicity of the alternating group, 149
Skew field, 36
Soluble group, 146
Solution by radicals, 172
Squaring of the circle, 186
Steinitz' replacement theorem, 101
Sturm's
chain, 220
theorem, 220
Subfield, 91
Subgroup, 18
admissible, 138
characteristic, 139
conjugate, 26
distinguished (invariant), 24
self-conjugate, 24
Subring, 49
Subset, 1
proper, 2
Substitution, 153
Sum
composite, 15
of squares, 225, 233
of two ideals, 55
of two numbers, 4
Symbolic adjunction, 99
Symmetric, 9
function, 78
group, 13
System of double composition, 32

Theorem
Abel's, 177
Lüroth's, 198
on the least upper bound, 216
on the primitive element, 126
Rolle's, 222
Topologically isomorphic, 244
Totally positive, 234
Trace, 128
reduced, 133
Transcendence, degree of, 201
Transcendental, 95
extension, 95
Transformation, 12
identical, 13
inverse, 13
Transformation
group, 13
Transitive, 9, 150
Transitivity set, 150
Transposition, 19
Trisection of an angle, 184

Uncountable, 8
Union, 2
Uniqueness
of prime factorization, 60
theorem on symmetric functions, 81
Unit, 59
ideal, 49
Unity operator, 139
Upper bound, 211

Valuation, 235
Archimedean, 237
discrete, 239
equivalent, 243
exponential, 238
non-Archimedean, 237
non-disorete, 239
p-adic, 236
Valuation ring, 239
Variable, 47
Vector space, 42

Weierstrass' Nullstellensatz, 218
Weight of a polynomial, 78
Wilson's theorem, 118

Zero, 6, 15
of a polynomial, 64
of real functions, 218
Zero divisor, 34